

# Information Networks with in-Block Memory

Gerhard Kramer

**Abstract**—An information network with in-block memory (NiBM) is a generalization of a discrete memoryless network (DMN) where blocks of symbols may have memory inside each block. A cut-set bound is developed for NiBMs that unifies, strengthens, and generalizes existing cut bounds. The bound gives finite-letter capacity expressions for several classes of networks including point-to-point channels with iBM, and certain multiaccess, broadcast, and relay channels with iBM. Cardinality bounds on the random coding alphabets are developed that improve on existing bounds for channels with action-dependent state available causally at the encoder and for relays without delay. Finally, digital network coding is shown to achieve rates within a limited gap of the new cut-set bound for linear, additive, Gaussian noise channels, symmetric power constraints, and a multicast session.

**Index Terms**—capacity, feedback, information, relay channels, networks

## I. INTRODUCTION

An information network with in-block memory (NiBM) is an extension of a discrete memoryless network (DMN). Recall that a DMN with  $K$  nodes has each node  $k$ ,  $k = 1, 2, \dots, K$ , dealing with four types of random variables [1].

- **Messages**  $W_{km}$ ,  $m = 1, 2, \dots, M_k$ , that have entropy  $H(W_{km}) = B_{km}$  bits where  $M_k$  is the number of messages at node  $k$ . The rate of message  $W_{km}$  is thus  $R_{km} = B_{km}/n$  bits per channel use. The  $\{W_{km}\}$  are mutually statistically independent for all  $m$  and  $k$ .
- **Channel inputs**  $X_{k,i}$ ,  $i = 1, 2, \dots, n$ , with alphabet  $\mathcal{X}_k$ . We interpret  $i$  as a time index but it could alternatively represent frequency or space, for example.
- **Channel outputs**  $Y_{k,i}$ ,  $i = 1, 2, \dots, n$ , with alphabet  $\mathcal{Y}_k$ .
- **Message estimates**  $\hat{W}_{\ell m}^{(k)}$ ,  $\ell m \in \mathcal{D}(k)$ , where  $\mathcal{D}(k)$  is a *decoding index set* whose elements are selected pairs  $\ell m$ ,  $\ell \neq k$ , of message indices from other nodes.

Let  $\mathcal{K} = \{1, 2, \dots, K\}$  be the set of nodes; let  $\mathcal{E}(k) = \{k1, k2, \dots, kM_k\}$  be the *encoding index set* of node  $k$ ; let  $Y_k^i = Y_{k,1}Y_{k,2} \dots Y_{k,i}$ ; let  $r(x, y)$  be the remainder when  $x$  is divided by  $y$ . For a set  $\mathcal{S} \subseteq \mathcal{K}$  we write  $\mathcal{E}(\mathcal{S}) = \cup_{k \in \mathcal{S}} \mathcal{E}(k)$  and  $X_{\mathcal{S},i} = \{X_{k,i} : k \in \mathcal{S}\}$ . For a set  $\tilde{\mathcal{S}}$  of integer pairs  $km$  we write  $W_{\tilde{\mathcal{S}}} = \{W_{km} : km \in \tilde{\mathcal{S}}\}$ . The relationships between the random variables are as follows.

Date of current version March 8, 2013. G. Kramer was supported by an Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research, and by NSF Grant CCF-09-05235. The material in this paper was presented in part at the First Workshop on Information Theory and Coding for Cooperative Networks, Paris, France, March 2012, the First Munich Workshop on Bidirectional Communication and Directed Information, Munich, Germany, May 2012, and the IEEE Information Theory Workshop, Lausanne, Switzerland, Sep. 2012.

The author is with the Institute for Communications Engineering, Technische Universität München, D-80333 Munich, Germany (email: gerhard.kramer@tum.de).

- Node  $k$  chooses functions  $\mathbf{a}_{k,i}$ ,  $i = 1, 2, \dots, n$ , such that

$$X_{k,i} = \mathbf{a}_{k,i}(W_{\mathcal{E}(k)}, Y_k^{i-1}). \quad (1)$$

For a finite alphabet  $\mathcal{Y}_k$  one may interpret  $\mathbf{a}_k^n(w_{\mathcal{E}(k)}, \cdot)$  as a *code function* (or *code tree* or *adaptive code word* or *strategy*) for the messages  $w_{\mathcal{E}(k)}$  (see [2, Sec. 15] and [3, Ch. 9]). We write  $\mathbf{a}_k^n(W_{\mathcal{E}(k)}, \cdot)$  as  $\mathbf{A}_k^n(W_{\mathcal{E}(k)}, \cdot)$  to emphasize that  $\mathbf{A}_k^n$  is a random function.

- Node  $k$  puts out

$$\hat{W}_{\mathcal{D}(k)}^{(k)} = d_k(W_{\mathcal{E}(k)}, Y_k^n) \quad (2)$$

for some decoding function  $d_k$ .

- A DMN channel is memoryless. A NiBM, however, may have in-block memory of length  $L$  in the sense that at time  $i$  node  $k$  sees

$$Y_{k,i} = f_{k,t(i)+1}(X_{\mathcal{K},i-t(i)}, \dots, X_{\mathcal{K},i}, Z_{\lceil i/L \rceil}) \quad (3)$$

$k = 1, 2, \dots, K$ , where  $t(i) = r(i-1, L)$ , and where the  $Z_i$ ,  $i = 1, 2, \dots, \lceil n/L \rceil$ , are statistically independent realizations of a random variable  $Z$  with alphabet  $\mathcal{Z}$ .

The channel functions  $f_{k,i}(\cdot)$ ,  $i = 1, 2, \dots, L$ , may be different and the alphabets  $\mathcal{X}_{k,i}$  and  $\mathcal{Y}_{k,i}$ ,  $i = 1, 2, \dots, L$ , may be different also. The notation  $\mathcal{X}_k^L$  means  $\mathcal{X}_{k,1} \times \mathcal{X}_{k,2} \times \dots \times \mathcal{X}_{k,L}$ .

*Example 1:* Consider a two-way channel with iBM of length  $L = 2$ . The channel puts out

- $Y_{k,1} = f_{k,1}(X_{1,1}, X_{2,1}, Z_1)$
- $Y_{k,2} = f_{k,2}(X_{1,1}, X_{1,2}, X_{2,1}, X_{2,2}, Z_1)$
- $Y_{k,3} = f_{k,1}(X_{1,3}, X_{2,3}, Z_2)$
- $Y_{k,4} = f_{k,2}(X_{1,3}, X_{1,4}, X_{2,3}, X_{2,4}, Z_2)$

for  $k = 1, 2$  and  $n = 4$ . A functional dependence graph (FDG) for this case is shown in Fig. 1 where the nodes  $W_1, W_2, Z_1, Z_2$  with hollow circles represent mutually statistically independent random variables [1], [4].

This paper develops information theory for NiBMs. Our main goal is to show that NiBMs are useful because much existing theory for DMNs extends naturally to NiBMs. Furthermore, NiBMs let us unify, strengthen, and generalize theory for several classes of networks, in particular for relay networks with delays. We believe that the framework of NiBMs is easier to understand than that of such specialized networks.

The document is organized as follows. Section II defines the capacity region of a NiBM and introduces notation. Section III states our main technical result: a cut-set bound on reliable communication rates. Sections IV and V apply the bound to point-to-point and multiuser channels. We derive a few new capacity theorems and cardinality bounds on random variables. Section VI extends these approaches to relay networks. Several proofs are developed in the Appendices.

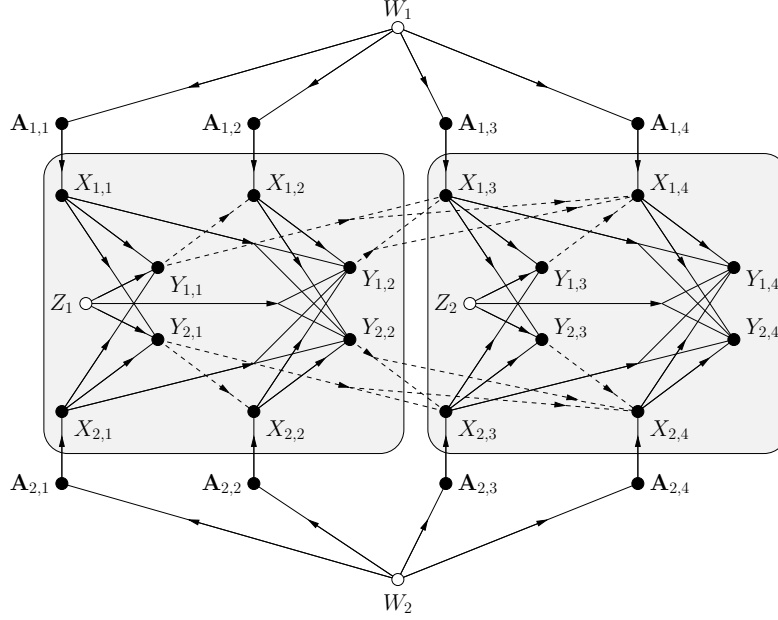


Fig. 1. FDG for a two-way channel with iBM of length  $L = 2$  and  $n = 4$  channel uses. The message estimates  $\hat{W}_1$  and  $\hat{W}_2$  are not shown. The two blocks of channel inputs and outputs are shaded and the functional dependence due to the received symbols is drawn with dashed lines. The code functions  $\mathbf{A}_1^n$  and  $\mathbf{A}_2^n$  are statistically independent.

## II. PRELIMINARIES

### A. Capacity

The *capacity region* of a NiBM is the closure of the set of rate-tuples  $(R_{km} : 1 \leq k \leq K, 1 \leq m \leq M_k)$  such that for any positive  $\epsilon$  there is an  $n$  and code functions and decoders for which the error probability

$$P_e = \Pr \left[ \bigcup_k \bigcup_{\ell m \in \mathcal{D}(k)} \{\hat{W}_{\ell m}^{(k)} \neq W_{\ell m}\} \right] \quad (4)$$

is at most  $\epsilon$ .

### B. Causal Conditioning and Directed Information

We use notation from [1] for causal conditioning and directed information. The probability of  $x^L$  causally conditioned on  $y^L$  and conditioned on  $a$  is defined as

$$P(x^L \| y^L) = \prod_{i=1}^L P(x_i | x^{i-1} y^i) \quad (5)$$

$$P(x^L \| y^L | a) = \prod_{i=1}^L P(x_i | x^{i-1} y^i a). \quad (6)$$

As done here, we will drop subscripts on probability distributions if the argument is the lowercase version of the random variable. Causally-conditioned entropy is defined as

$$H(X^L \| Y^L) = \sum_{i=1}^L H(X_i | X^{i-1} Y^i) \quad (7)$$

$$H(X^L \| Y^L | A) = \sum_{i=1}^L H(X_i | X^{i-1} Y^i A). \quad (8)$$

Directed information is written as

$$I(X^L \rightarrow Y^L) = H(Y^L) - H(Y^L \| X^L) \quad (9)$$

$$I(X^L \rightarrow Y^L \| Z^L) = H(Y^L \| Z^L) - H(Y^L \| X^L, Z^L) \quad (10)$$

$$I(X^L \rightarrow Y^L \| Z^L | A) \quad (11)$$

$$= H(Y^L \| Z^L | A) - H(Y^L \| X^L, Z^L | A). \quad (12)$$

### C. Further Notation

The functional dependence (1) is expressed as  $1(x_{k,i} | \mathbf{a}_k^i, y_k^{i-1})$  in place of  $P(x_{k,i} | \mathbf{a}_k^i, y_k^{i-1})$ . We similarly write  $1(x_k^L | \mathbf{a}_k^L, 0y_k^{L-1})$  in place of  $P(x_k^L | \mathbf{a}_k^L, 0y_k^{L-1})$ . It will be convenient to split symbol strings into blocks of length  $L$ . We use the notation

$$\mathbf{a}_{k,i}^L = \mathbf{a}_{k,i(m-1)+1} \mathbf{a}_{k,i(m-1)+2} \cdots \mathbf{a}_{k,i(m-1)+L}$$

$$x_{k,i}^L = x_{k,i(m-1)+1} x_{k,i(m-1)+2} \cdots x_{k,i(m-1)+L}$$

$$y_{k,i}^L = y_{k,i(m-1)+1} y_{k,i(m-1)+2} \cdots y_{k,i(m-1)+L}.$$

We write  $\text{supp}(P_X)$  for the support set of  $P_X(\cdot)$ . We write the binary entropy function as  $H_2(\cdot)$  and differential entropy as  $h(\cdot)$ . Logarithms are taken to the base 2.

### D. Channel Distribution

We have defined the channel using the *function* (3). It will be convenient to alternatively define the channel by a *probability distribution*. Consider  $P(\mathbf{a}_{\mathcal{K}}^n, x_{\mathcal{K}}^n, y_{\mathcal{K}}^n)$  that factors as

$$\left[ \prod_{k=1}^K P(\mathbf{a}_k^n) 1(x_k^n | \mathbf{a}_k^n, 0y_k^{n-1}) \right] P(y_{\mathcal{K}}^n | x_{\mathcal{K}}^n). \quad (13)$$

The  $P(y_{\mathcal{K}}^n | x_{\mathcal{K}}^n)$  further factors into  $m = \lceil n/L \rceil$  blocks as

$$\left[ \prod_{i=1}^{m-1} P_{Y_{\mathcal{K}}^L \| X_{\mathcal{K}}^L}(y_{\mathcal{K},i}^L | x_{\mathcal{K},i}^L) \right] P_{Y_{\mathcal{K}}^L \| X_{\mathcal{K}}^L}(y_{\mathcal{K},m}^L | x_{\mathcal{K},m}^L) \quad (14)$$

where the last block has length  $L' = n - (m-1)L$ . We focus on the case  $n = mL$  so that  $L' = L$  and all blocks have length  $L$ . The expression (14) means that we may define the NiBM channel by using the block-invariant distribution  $P(y_k^L \| x_k^L)$  rather than by using  $Z$  and the functions in (3).

### E. Linear Channels

We consider several examples where the channel alphabets are the field  $\mathbb{F}$ . We write the channel inputs and outputs as vectors  $\underline{X}_k = [X_{k,1} \dots X_{k,L}]^T$  and  $\underline{Y}_k = [Y_{1,k} \dots Y_{k,L}]^T$ , respectively. For instance, a *scalar*, *linear*, and *additive-noise* channel has

$$\underline{Y}_k = \left[ \sum_{j \neq k} \mathbf{G}_{kj} \underline{X}_j \right] + \underline{Z}_k \quad (15)$$

where the  $\mathbf{G}_{kj}$  are  $L \times L$  lower-triangular matrices and the  $\underline{Z}_k$  are random vectors such that  $\underline{Z}^K$  is independent of  $\underline{X}^K$ . We write the covariance matrix of a random vector  $\underline{X}$  as  $\mathbf{Q}_{\underline{X}}$  and its determinant as  $|\mathbf{Q}_{\underline{X}}|$ .

## III. CUT-SET BOUND

We develop a cut-set bound for NiBM that generalizes the classic cut-set bound for DMNs. Consider a set  $\mathcal{S}$  of nodes and let  $\mathcal{S}^c$  be the complement of  $\mathcal{S}$  in  $\mathcal{K} = \{1, 2, \dots, K\}$ . We say that  $(\mathcal{S}, \mathcal{S}^c)$  is a *cut separating* a message  $W_{km}$  and its estimate  $\hat{W}_{km}^{(\ell)}$  if  $k \in \mathcal{S}$  and  $\ell \in \mathcal{S}^c$ . Let  $\mathcal{M}(\mathcal{S})$  be the set of indexes (which are integer pairs  $km$ ) of those messages separated from one of their estimates by the cut  $(\mathcal{S}, \mathcal{S}^c)$ , and let  $R_{\mathcal{M}(\mathcal{S})}$  be the sum of the rates of these messages.

There is a subtlety in that the NiBM can have high mutual information at the start of each block and low mutual information at the end of each block. This could mean, e.g., that using the channel 1 time is better than using it a large number of times. To avoid such issues, we require that the channel is used  $n = mL$  times for a positive integer  $m$ . Alternatively, we could require that  $n$  be much larger than  $L$ . We have the following result that we prove in Appendix A.

*Theorem 1:* The capacity region  $\mathcal{C}$  of a NiBM of length  $L$  that is used a multiple of  $L$  times satisfies

$$\mathcal{C} \subseteq \bigcup_{P_{\mathbf{A}_{\mathcal{K}}^L}} \bigcap_{\mathcal{S} \subset \mathcal{K}} \mathcal{R}(P_{\mathbf{A}_{\mathcal{K}}^L}, \mathcal{S}) \quad (16)$$

where  $\mathcal{R}(P_{\mathbf{A}_{\mathcal{K}}^L}, \mathcal{S})$  is the set of non-negative rate-tuples satisfying

$$R_{\mathcal{M}(\mathcal{S})} \leq I(\mathbf{A}_{\mathcal{S}}^L; Y_{\mathcal{S}^c}^L | \mathbf{A}_{\mathcal{S}^c}^L) / L. \quad (17)$$

The joint probability distribution  $P(\mathbf{a}_{\mathcal{K}}^L, x_{\mathcal{K}}^L, y_{\mathcal{K}}^L)$  factors as

$$P(\mathbf{a}_{\mathcal{K}}^L) \left[ \prod_{k=1}^K 1(x_k^L \| \mathbf{a}_k^L, 0y_k^{L-1}) \right] P(y_{\mathcal{K}}^L \| x_{\mathcal{K}}^L). \quad (18)$$

*Remark 1:* The code functions in Theorem 1 are statistically *dependent*. This is different than in Sec. I where the code functions are independent (see Fig. 1 and (13)). Similarly, Shannon's outer bound for the two-way channel [2, Eq. (36)] and the classic cut-set bound for DMNs [1], [4, Ch. 10], [5,

Sec. 15.10], [6, p. 477] have statistically *dependent* inputs (see Sec. III-B).

*Remark 2:* The  $1(x_k^L \| \mathbf{a}_k^L, 0y_k^{L-1})$ ,  $k = 1, 2, \dots, K$ , are fixed functions and  $P(y_{\mathcal{K}}^L \| x_{\mathcal{K}}^L)$  is fixed by the channel.

*Remark 3:* Fixing  $P(y_{\mathcal{K}}^L \| x_{\mathcal{K}}^L)$  fixes  $P(y_{\mathcal{K}}^L | \mathbf{a}_{\mathcal{K}}^L)$ . We may thus view the channel as being  $P(y_{\mathcal{K}}^L | \mathbf{a}_{\mathcal{K}}^L)$  for the purposes of deriving achievable rates and computing the cut-set bound.

*Remark 4:*  $I(\mathbf{A}_{\mathcal{S}}^L; Y_{\mathcal{S}^c}^L | \mathbf{A}_{\mathcal{S}^c}^L)$  is concave in  $P_{\mathbf{A}_{\mathcal{K}}^L}$ . This result follows by the concavity of  $I(A; B | C = c)$  in  $P_{A|C=c}$  when  $P_{B|A=C=c}$  is held fixed, and because  $P(y_{\mathcal{K}}^L | \mathbf{a}_{\mathcal{K}}^L)$  is fixed.

*Remark 5:* The  $\mathbf{A}_{\mathcal{K}}^L$  are *not* “auxiliary” random variables, i.e., they are explicit components of the communication problem just like the channel inputs  $X_k^L$ . Moreover, the cardinalities  $|\mathcal{A}_{\mathcal{K}}^L|$  are automatically bounded by the channel alphabets.

*Remark 6:* Average per-letter cost constraints can be dealt with in the usual way (see Remark 29). More precisely, if we have  $J$  cost functions  $s_j(\cdot)$  and constraints

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[s_j(X_{\mathcal{K},i}, Y_{\mathcal{K},i})] \leq S_j, \quad j = 1, 2, \dots, J \quad (19)$$

then one may add the requirement that the union in (16) is over distributions (18) that satisfy

$$\frac{1}{L} \sum_{i=1}^L \mathbb{E}[s_j(X_{\mathcal{K},i}, Y_{\mathcal{K},i})] \leq S_j, \quad j = 1, 2, \dots, J. \quad (20)$$

One may treat average per-block cost constraints similarly.

*Remark 7:* A bound in [7, Thm. 2] and [8, Thm. 2] is similar to (17) for the special cases of relay networks with delays and causal relay networks without messages at the causal relays. We discuss these bounds in more detail in Remark 26 below.

### A. Weakened Bounds

The bound (17) may be weakened as follows:

$$\begin{aligned} & I(\mathbf{A}_{\mathcal{S}}^L; Y_{\mathcal{S}^c}^L | \mathbf{A}_{\mathcal{S}^c}^L) \\ & \stackrel{(a)}{=} \sum_{i=1}^L H(Y_{\mathcal{S}^c,i} | Y_{\mathcal{S}^c}^{i-1} \mathbf{A}_{\mathcal{S}^c}^L) - H(Y_{\mathcal{S}^c,i} | Y_{\mathcal{S}^c}^{i-1} \mathbf{A}_{\mathcal{K}}^i) \\ & \leq \sum_{i=1}^L H(Y_{\mathcal{S}^c,i} | Y_{\mathcal{S}^c}^{i-1} \mathbf{A}_{\mathcal{S}^c}^i) - H(Y_{\mathcal{S}^c,i} | Y_{\mathcal{S}^c}^{i-1} \mathbf{A}_{\mathcal{K}}^i) \\ & = I(\mathbf{A}_{\mathcal{S}}^L \rightarrow Y_{\mathcal{S}^c}^L | \mathbf{A}_{\mathcal{S}^c}^L) \end{aligned} \quad (21)$$

where (a) follows by the chain rule for entropy and because

$$(\mathbf{A}_{\mathcal{K},i+1} \dots \mathbf{A}_{\mathcal{K},L}) - \mathbf{A}_{\mathcal{K}}^i Y_{\mathcal{S}^c}^{i-1} - Y_{\mathcal{S}^c,i}$$

forms a Markov chain. The bound (21) may be further weakened to replace code functions with channel inputs and outputs:

$$\begin{aligned} & I(\mathbf{A}_{\mathcal{S}}^L \rightarrow Y_{\mathcal{S}^c}^L | \mathbf{A}_{\mathcal{S}^c}^L) \\ & \stackrel{(a)}{\leq} \sum_{i=1}^L H(Y_{\mathcal{S}^c,i} | Y_{\mathcal{S}^c}^{i-1} X_{\mathcal{S}^c}^i) - H(Y_{\mathcal{S}^c,i} | Y_{\mathcal{K}}^{i-1} X_{\mathcal{K}}^i \mathbf{A}_{\mathcal{K}}^i) \\ & \stackrel{(b)}{=} I(X_{\mathcal{S}}^L, 0Y_{\mathcal{S}}^{L-1} \rightarrow Y_{\mathcal{S}^c}^L | X_{\mathcal{S}^c}^L) \end{aligned} \quad (22)$$

where (a) follows because  $Y_k^{i-1} \mathbf{A}_k^i$  defines  $X_k^i$  and because conditioning cannot increase entropy. Step (b) follows because (18) ensures that the chain  $\mathbf{A}_K^L - Y_K^{i-1} X_K^i - Y_{K,i}$  is Markov.

*Remark 8:* The FDG of a NiBM has statistically independent code functions, see Fig. 1. We thus have

$$H(Y_{S^c,i} | Y_{S^c}^{i-1} X_{S^c}^i \mathbf{A}_{S^c}^L) = H(Y_{S^c,i} | Y_{S^c}^{i-1} X_{S^c}^i). \quad (23)$$

However, the identity (23) may not be valid when considering dependent code functions such as in Theorem 1.

*Remark 9:* The cut-set bound with the normalized (22) in place of the right-hand side of (17) was derived in [8, Thm. 1] for causal relay networks and in [9, Thm. 1] for generalized networks. The authors of [8], [9] restrict attention to multiple unicast sessions as in [5, Sec. 15.10]. Combining Theorem 1 and (22) extends the bounds to multiple multicast sessions. We discuss these bounds in more detail in Sec. VI-C.

*Example 2:* Consider *additive* noise channels with

$$Y_{k,i} = f_{k,i}(X_K^i) + Z_{k,i} \quad (24)$$

for  $i = 1, 2, \dots, L$ ,  $k = 1, 2, \dots, K$ , where  $Y_{k,i}$ ,  $Z_{k,i}$ , and  $f_{k,i}(X_K^i)$  take on values in the field  $\mathbb{F}$ . The noise variables  $Z_K^L$  are independent of  $X_K^L$ . The bound (22) is

$$\begin{aligned} I(\mathbf{A}_S^L; Y_{S^c}^L | \mathbf{A}_{S^c}^L) &\leq I(X_S^L, 0Y_S^{L-1} \rightarrow Y_{S^c}^L | X_{S^c}^L) \\ &= H(Y_{S^c}^L | X_{S^c}^L) - H(Z_{S^c}^L | 0Z_S^{L-1}). \end{aligned} \quad (25)$$

Since  $H(Z_{S^c}^L | 0Z_S^{L-1})$  is fixed by the channel, the cut-set bound with the normalized (25) in place of the right-hand side of (17) is a maximum (conditional) entropy problem.

*Example 3:* A special case of (24) is a *deterministic* NiBM for which the noise is a constant and

$$I(\mathbf{A}_S^L; Y_{S^c}^L | \mathbf{A}_{S^c}^L) \leq H(Y_{S^c}^L | X_{S^c}^L). \quad (26)$$

## B. DMNs

Suppose the NiBM is a DMN. We have  $L = 1$  and recover the classic cut-set bound. Alternatively, we may view the DMN as an NiBM of length  $L$  and with

$$P(y_K^L | x_K^L) = \prod_{i=1}^L P_{Y_{K,i} | X_{K,i}}(y_{K,i} | x_{K,i}). \quad (27)$$

The weakened bound (22) becomes

$$\begin{aligned} I(X_S^L, 0Y_S^{L-1} \rightarrow Y_{S^c}^L | X_{S^c}^L) \\ &= \sum_{i=1}^L H(Y_{S^c,i} | X_{S^c}^i Y_{S^c}^{i-1}) - H(Y_{S^c,i} | X_{K,i}) \\ &\leq \sum_{i=1}^L I(X_{S,i}; Y_{S^c,i} | X_{S^c,i}). \end{aligned} \quad (28)$$

If we choose the code functions as code words and

$$P(x_K^L) = \prod_{i=1}^L P(x_{K,i}) \quad (29)$$

then we achieve equality in (28). We recover the classic cut-set bound by choosing  $P(x_{K,i}) = P_{X_K}(x_{K,i})$  for all  $i$ .

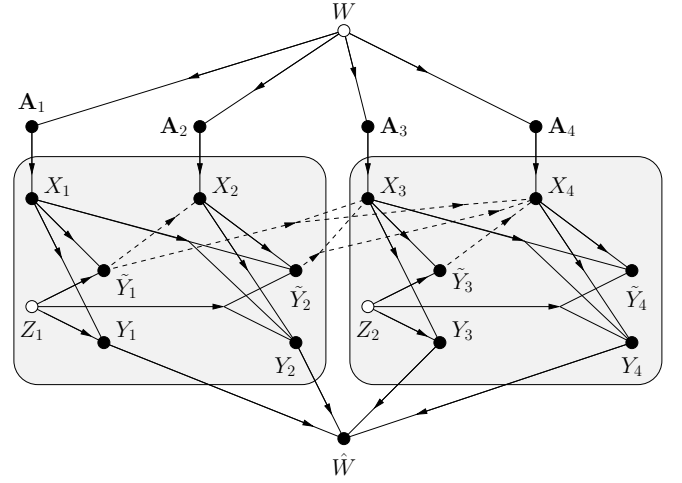


Fig. 2. FDG for a point-to-point channel with iBM of length  $L = 2$  and  $n = 4$  channel uses.

*Remark 10:* Consider a DMN that is time varying in blocks of length  $L$ , i.e., we have an NiBM of length  $L$  and

$$P(y_K^L | x_K^L) = \prod_{i=1}^L P_{Y_{K,i} | X_{K,i}}(y_{K,i} | x_{K,i}) \quad (30)$$

The cut-set bound of Theorem 1 may be computed with independent inputs as in (29).

## IV. POINT-TO-POINT CHANNELS

Consider a point-to-point channel with input  $X^L$  taking on values in  $\mathcal{X}^L$ , receiver output  $Y^L$  taking on values in  $\mathcal{Y}^L$ , and feedback  $\tilde{Y}^L$  taking on values in  $\tilde{\mathcal{Y}}^L$ . A FDG for  $L = 2$  and  $n = 4$  is shown in Fig. 2.

*Theorem 2:* The capacity of a point-to-point channel with iBM of length  $L$  is

$$C = \max_{P_{\mathbf{A}^L}} I(\mathbf{A}^L; Y^L) / L \quad (31)$$

where  $P(\mathbf{a}^L, y^L, \tilde{y}^L)$  factors as

$$P(\mathbf{a}^L) 1(x^L | \mathbf{a}^L, 0\tilde{y}^{L-1}) P(y^L, \tilde{y}^L | x^L). \quad (32)$$

*Proof:* Achievability follows by random coding with a maximizing  $P_{\mathbf{A}^L}$ . The converse follows by Theorem 1. ■

*Remark 11:* The distribution (32) gives

$$I(\mathbf{A}^L; Y^L) = I(\mathbf{A}^L \rightarrow Y^L). \quad (33)$$

*Remark 12:* The feedback  $\tilde{Y}^L$  can be noisy.

*Remark 13:* In-block feedback can increase  $C$  but across-block feedback does not increase  $C$ . This statement refines Shannon's classic theorem on feedback capacity [10, Thm. 6]. The transmitter can thus ignore  $\tilde{Y}_{iL}$  for all  $i$ , e.g., we could remove  $\tilde{Y}_2$  and  $\tilde{Y}_4$  in Fig. 2 without changing  $C$ .

*Remark 14:*  $I(\mathbf{A}^L; Y^L)$  is concave in  $P_{\mathbf{A}^L}$  and the Arimoto-Blahut algorithm [11], [12] can perform the maximization (31).

The cardinality  $|\mathcal{A}^L|$  is bounded by the channel alphabets (see Remark 5) and we have

$$|\mathcal{A}^L| = \prod_{i=1}^L |\mathcal{X}_i|^{|\tilde{\mathcal{Y}}^{i-1}|}. \quad (34)$$

Unfortunately, (34) means that  $|\mathcal{A}^L|$  grows doubly exponentially in  $L$  if the alphabet sizes are similar for all  $i$ . We prove the following Theorem by using classic results [13, p. 96], [14, p. 310] on bounding set sizes.

**Theorem 3:** The maximum in Theorem 2 is achieved by a  $P_{\mathbf{A}^L}$  for which  $|\text{supp}(P_{\mathbf{A}^L})|$  is at most

$$\min \left( |\mathcal{Y}^L|, |\mathcal{X}_1| + \sum_{i=2}^L |\mathcal{X}^{i-1}| \cdot |\tilde{\mathcal{Y}}^{i-1}| \cdot (|\mathcal{X}_i| - 1) \right). \quad (35)$$

*Proof:* See Appendix B. ■

**Remark 15:** Theorem 3 ensures that  $|\text{supp}(P_{\mathbf{A}^L})|$  must grow only exponentially, and not doubly exponentially, in  $L$ . Of course, one must still determine  $\text{supp}(P_{\mathbf{A}^L})$  which can be a high-complexity search problem for even small  $L$ .

**Example 4:** Consider a channel with  $L = 2$  that has binary symmetric channels (BSCs) with

$$Y_1 = X_1 \oplus Z_1, \quad \tilde{Y}_1 = Z_1, \quad Y_2 = X_2 \oplus Z_1 \oplus \tilde{Z}_2. \quad (36)$$

The bits  $Z_1$  and  $\tilde{Z}_2$  are independent and  $P_{Z_1}(1) = \epsilon_1$  and  $P_{\tilde{Z}_2}(1) = \epsilon_2$ . This is an additive noise channel of the form (24) and for  $\mathcal{S} = \{1\}$  we compute (see (25))

$$\begin{aligned} H(Z_{\mathcal{S}^c}^L \| 0 Z_{\mathcal{S}}^{L-1}) &= H(Z_1) + H(Z_1 \oplus \tilde{Z}_2 | Z_1) \\ &= H_2(\epsilon_1) + H_2(\epsilon_2). \end{aligned} \quad (37)$$

To compute the capacity, consider the steps

$$\begin{aligned} I(\mathbf{A}^2; Y^2) &= H(Y^2) - [H(Y_1 | X_1) + H(Y_2 | Y_1 X_1 \mathbf{A}_2)] \\ &\stackrel{(a)}{=} H(Y^2) - [H_2(\epsilon_1) + H_2(\epsilon_2)] \\ &\stackrel{(b)}{\leq} 2 - [H_2(\epsilon_1) + H_2(\epsilon_2)] \end{aligned} \quad (38)$$

where (a) follows because  $Y_1 X_1 \mathbf{A}_2$  determine  $Z_1 \tilde{Y}_1 X_2$ , and with equality in (b) if  $X_1$  and  $X_2$  are independent and uniformly distributed bits. The capacity is thus given by the right-hand side of (38).

For instance, we may transmit  $X_2 = X'_2 \oplus Z_1$  where  $X'_2$  is independent of  $X_1$ . We translate this strategy into a code function (here a code tree) distribution. We label  $\mathbf{A}^2$  as  $b, b_0 b_1$  by which we mean that  $X_1 = b$ ,  $X_2 = b_0$  if  $\tilde{Y}_1 = 0$ , and  $X_2 = b_1$  if  $\tilde{Y}_1 = 1$ . We choose

$$\begin{aligned} P_{\mathbf{A}^2}(0, 00) &= P_{\mathbf{A}^2}(0, 11) = P_{\mathbf{A}^2}(1, 00) = P_{\mathbf{A}^2}(1, 11) = 0 \\ P_{\mathbf{A}^2}(0, 01) &= P_{\mathbf{A}^2}(0, 10) = P_{\mathbf{A}^2}(1, 01) = P_{\mathbf{A}^2}(1, 10) = 1/4 \end{aligned}$$

and achieve capacity with four code trees, as predicted by Theorem 3.

**Example 5:** We show the deficiencies of the weakened bound based on (22). Suppose the channel is

$$Y_1 = X_1 \oplus \tilde{Z}_1 \oplus Z_2, \quad \tilde{Y}_1 = \tilde{Z}_1, \quad Y_2 = Z_2 \quad (39)$$

where  $\tilde{Z}_1$  and  $Z_2$  are independent with  $P_{\tilde{Z}_1}(1) = \epsilon_1$  and  $P_{Z_2}(1) = \epsilon_2$ . We achieve

$$C = (1 - H_2(\epsilon_1))/2$$

by having the receiver compute  $Y_1 \oplus Y_2 = X_1 \oplus \tilde{Z}_1$ . In other words, we achieve capacity with uniform  $X_1$  and so we require only two code trees  $\mathbf{a}^2 = 0, 00$  and  $\mathbf{a}^2 = 1, 00$ .

For the weakened bound (25), observe that (39) has the form (24). Defining  $\epsilon_1 * \epsilon_2 = \epsilon_1(1 - \epsilon_2) + (1 - \epsilon_1)\epsilon_2$  and  $\mathcal{S} = \{1\}$  we compute

$$\begin{aligned} H(Z_{\mathcal{S}^c}^L \| 0 Z_{\mathcal{S}}^{L-1}) &= H(\tilde{Z}_1 \oplus Z_2) + H(Z_2 | \tilde{Z}_1 \oplus Z_2, \tilde{Z}_1) \\ &= H_2(\epsilon_1 * \epsilon_2). \end{aligned} \quad (40)$$

The weakened bound (25) is therefore

$$\begin{aligned} 2C &\leq \max_{P_{X^2 \| 0 Y_1}} H(Y^2) - H_2(\epsilon_1 * \epsilon_2) \\ &= 1 + H_2(\epsilon_2) - H_2(\epsilon_1 * \epsilon_2) \end{aligned} \quad (41)$$

with equality if  $X_1$  is uniform. This bound is loose in general, e.g., if  $\epsilon_1 = 1/2$  then  $C = 0$  but (41) gives  $C \leq H_2(\epsilon_2)/2$ .

#### A. Noise-Free Feedback

The feedback is *noise-free* if  $\tilde{Y}^L$  is a causal function of  $X^L$  and  $Y^L$ , i.e., if  $\tilde{Y}_i = f_i(X^i, Y^i)$  for  $i = 1, 2, \dots, L$ . For instance, the channel (36) has noise-free feedback. The expression (31) simplifies to

$$C = \max_{P_{X^L \| 0 Y^{L-1}}} I(X^L \rightarrow Y^L)/L. \quad (42)$$

**Example 6:** Consider the additive noise channel (15) with

$$\underline{Y} = \mathbf{G}\underline{X} + \underline{Z} \quad (43)$$

where  $\underline{Z}$  is independent of  $\underline{X}$ . We compute

$$H(Y^L \| X^L) = H(Z^L) \quad (44)$$

so that computing (42) reduces to maximizing  $H(Y^L)$ . For instance, for modulo-additive channels the maximizing  $X^L$  will be uniformly distributed over  $\mathbb{F}^L$ , and for additive Gaussian noise (AGN) channels with transmit power constraints the maximizing  $X^L$  will be Gaussian.

#### B. Block Fading Channels

Channels with *block interference* [15] or *block fading* [16] have a *state*  $S$  that is memoryless across blocks of length  $L$  and whose realization  $S = s$  specifies a memoryless channel in each block. In other words, when  $S = s$  we have

$$P(y^L, \tilde{y}^L \| x^L | s) = \prod_{i=1}^L P_{Y\tilde{Y}|X S}(y_i, \tilde{y}_i | x_i, s). \quad (45)$$

We may view such channels as NiBMs for which  $Z = SN^L$ , i.e.,  $Z$  includes the state  $S$  and a noise string  $N^L$  where the  $N_i$ ,  $i = 1, 2, \dots, L$ , are statistically independent and identically distributed. Equation (3) thus becomes

$$Y_i = f_{t(i)+1}(X_{i-t(i)}, \dots, X_i, S_{\lceil i/L \rceil} N_i) \quad (46)$$

$$\tilde{Y}_i = \tilde{f}_{t(i)+1}(X_{i-t(i)}, \dots, X_i, S_{\lceil i/L \rceil} N_i) \quad (47)$$

for  $i = 1, 2, \dots, n$ .

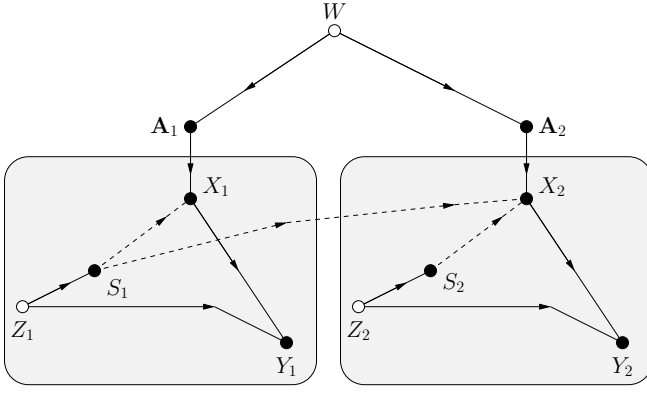


Fig. 3. FDG for a channel with state known causally at the encoder. The NiBM has  $L = 2$ . The message estimate  $\hat{W}$  is not shown.

### C. Channels with State Known Causally at the Encoder

Shannon's channel [17] with state known causally at the encoder is a point-to-point channel with iBM of length  $L = 2$ . To see this, choose  $\mathcal{X}_1 = \mathcal{Y}_1 = \{0\}$  and  $\tilde{Y}_1 = S$  where  $S$  is the state. A FDG is shown in Fig. 3 where we have renamed the random variables  $\mathbf{A}$ ,  $S$ ,  $X$ ,  $Y$ , and  $Z$ . In the standard model, the noise  $Z$  has two components  $SZ'$  where  $Z'$  is independent of  $S$  and  $Y = f(X, S, Z')$  for some function  $f(\cdot)$ .

The capacity is given by Theorem 2 and normalizing by a factor of  $L = 2$ . Moreover, the alphabet size of  $\mathbf{A}$  is  $|\mathcal{X}|^{|\mathcal{S}|}$  but (35) tells us that

$$|\text{supp}(P_{\mathbf{A}})| \leq \min(|\mathcal{Y}|, 1 + |\mathcal{S}| \cdot (|\mathcal{X}| - 1)) \quad (48)$$

suffices. The  $|\mathcal{Y}|$  bound is due to Shannon [17] and the second bound was reported in [18, Thm. 1].

*Example 7:* Suppose that  $S = \mathcal{X} = \mathcal{Y} = \{0, 1\}$ ,  $P_S(0) = 1/2$ , and  $Y = X \oplus S$ . We label the branch-pairs  $\mathbf{A}$  as  $b_0 b_1$ , by which we mean that  $X = b_0$  if  $S = 0$  and  $X = b_1$  if  $S = 1$ . The capacity is clearly 1 and by choosing

$$\begin{aligned} P_{\mathbf{A}}(00) &= P_{\mathbf{A}}(11) = 0 \\ P_{\mathbf{A}}(01) &= P_{\mathbf{A}}(10) = 1/2 \end{aligned}$$

we attain  $I(\mathbf{A}; Y) = 1$  with two code trees, as predicted by Theorem 3. Note that  $I(X; Y) = 0$  is not the capacity. The weakened bound (22) happens to give  $C = I(XS; Y) = 1$ .

*Remark 16:* The above construction extends in an obvious way to show that any DMN with state(s) known causally at the encoder(s) is effectively an NiBM of length  $L = 2$ . The cut-set bound (16) thus applies to these problems.

### D. Channels with Action-Dependent State

Weissman's channel with action-dependent state lets the transmitter influence the state [19]. If the state is available causally at the encoder, then this model is a point-to-point channel with iBM of length  $L = 2$ . We treat the model shown in Fig. 4: at time  $i = 1$  the action  $B$  leads to the feedback state  $S$ , and at time  $i = 2$  the channel input and output is  $X$  and  $Y$ , respectively. Theorem 2 gives the capacity

$$2C = \max_{P_{\tilde{\mathbf{A}}\mathbf{A}}} I(\tilde{\mathbf{A}}\mathbf{A}; Y) = \max_{P_{\mathbf{B}\mathbf{A}}} I(\mathbf{B}\mathbf{A}; Y) \quad (49)$$

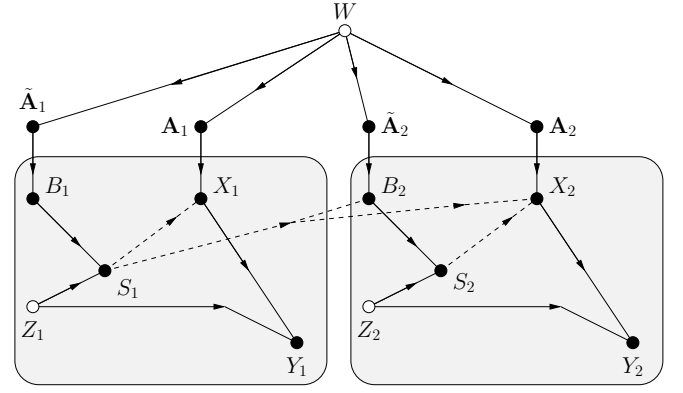


Fig. 4. FDG for a channel with action-dependent state known causally at the encoder. The NiBM has  $L = 2$  and the actions are  $B_1$  and  $B_2$ .

and Theorem 3 gives

$$|\text{supp}(P_{\tilde{\mathbf{A}}\mathbf{A}})| \leq \min(|\mathcal{Y}|, |\mathcal{B}| + |\mathcal{B}| |\mathcal{S}| (|\mathcal{X}| - 1)). \quad (50)$$

*Remark 17:* The expression (49) is the same as in [19, Thm. 2] because  $U$  plays the role of  $\tilde{\mathbf{A}}\mathbf{A}$ .

*Remark 18:* The constraint (50) is slightly stronger than that in [19, Thm. 2].

*Remark 19:* The model in Fig. 4 seems more general than in [19] because  $Z$  may influence both  $S$  and  $Y$ . However, the associations described in [19, p. 5405] show that the original model includes more problems than apparent at first glance. (See also comments in [19, Sec. VII].)

*Remark 20:* The model in Fig. 4 may seem different than in [19] because  $S$  may influence future actions as well as the present and future  $X$ . However, across-block feedback does not increase capacity (see Remark 13) so we may remove the  $S$ -to- $B$  functional dependence without affecting capacity. (See also comments in [19, Sec. VII] concerning feedback).

*Remark 21:* We may add functional dependence from  $B$  to  $Y$  without changing the capacity expression. Similar comments are made in [19, p. 5398 and Sec. VII].

*Example 8:* Consider a channel with a rewrite option [19, Sec. V.A] which means that the  $B$ -to- $S$  and  $X$ -to- $Y$  channels are effectively the same. At time  $i = 1$  the encoder "writes" on the  $B$ -to- $S$  channel. At time  $i = 2$ , if the encoder is happy with the outcome  $S$  then it sends a no-rewrite symbol  $N$  which means that  $Y = S$ . But if the encoder is unhappy with  $S$  then it "rewrites" a symbol on the  $X$ -to- $Y$  channel.

We have  $\mathcal{X} = \mathcal{B} \cup \{N\}$ ,  $\mathcal{S} = \mathcal{Y}$ , and the bound (50) is  $|\text{supp}(P_{\tilde{\mathbf{A}}\mathbf{A}})| \leq |\mathcal{Y}|$ . For example, suppose the  $B$ -to- $S$  channel is a BSC with crossover probability  $\delta$ ,  $0 \leq \delta \leq 1/2$  (see [19]). We label  $\tilde{\mathbf{A}}\mathbf{A}$  as  $b, b_0 b_1$  by which we mean that  $B = b$ ,  $X = b_0$  if  $S = 0$ , and  $X = b_1$  if  $S = 1$ . We have  $|\mathcal{Y}| = 2$  and achieve  $C = I(\tilde{\mathbf{A}}\mathbf{A}; Y) = 1 - H_2(\delta^2)$  by choosing

$$P_{\tilde{\mathbf{A}}\mathbf{A}}(0, N0) = P_{\tilde{\mathbf{A}}\mathbf{A}}(1, 1N) = 1/2.$$

We require only two code trees, as predicted by Theorem 3.

*Remark 22:* Multiple rewrites are modeled by increasing  $L$ .

## V. MULTIUSER CHANNELS

### A. Multiaccess Channels

Consider a two-user (three-terminal) MAC with iBM and with inputs  $X_1^L, X_2^L$ , and outputs  $Y^L, Y_1^L, Y_2^L$ . The FDG for  $L = 2$  and  $n = 4$  is the same as Fig. 1 except that the variables  $Y_i, i = 1, 2, 3, 4$ , are missing in Fig. 1. The cut-set bound of Theorem 1 is

$$\bigcup_{P_{\mathbf{A}_1^L \mathbf{A}_2^L}} \left\{ (R_1, R_2) : \begin{array}{l} 0 \leq R_1, 0 \leq R_2 \\ R_1 \leq I(\mathbf{A}_1^L; Y^L | \mathbf{A}_2^L) / L \\ R_2 \leq I(\mathbf{A}_2^L; Y^L | \mathbf{A}_1^L) / L \\ R_1 + R_2 \leq I(\mathbf{A}_1^L \mathbf{A}_2^L; Y^L) / L \end{array} \right\}. \quad (51)$$

If there is no feedback, then the cut-set bound can be strengthened in the usual way to

$$\bigcup \left\{ (R_1, R_2) : \begin{array}{l} 0 \leq R_1, 0 \leq R_2 \\ R_1 \leq I(X_1^L; Y^L | X_2^L T) / L \\ R_2 \leq I(X_2^L; Y^L | X_1^L T) / L \\ R_1 + R_2 \leq I(X_1^L X_2^L; Y^L | T) / L \end{array} \right\} \quad (52)$$

where the union is over distributions such that  $X_1^L - T - X_2^L$  forms a Markov chain ( $T$  is the usual time-sharing random variable). This modified cut-set bound is the capacity region without feedback. The result is not new, however, since the model is a special case of a classic MAC with vector alphabets.

*Remark 23:* MACs with state known causally at the encoders were treated in [22, Sec. IV]. As pointed out in Remark 16, such channels are NiBMs of length  $L = 2$ . For example, the outer bound of Theorem 3 in [22, Sec. IV] is the same as the cut-set bound of Theorem 1.

### B. Multiaccess Channels with Feedback

Several capacity results for DMNs generalize to problems with iBM. For example, consider Willems' result [20] that the Cover-Leung region [21] is  $\mathcal{C}$  for full feedback ( $Y_1 = Y_2 = Y$ ) and where one channel input, say  $X_1$ , is a function of  $Y$  and  $X_2$ . A natural generalization to MACs with iBM is to consider full feedback ( $Y_{1,i} = Y_{2,i} = Y_i$ ) and require  $X_{1,i} = f_i(X_2^i, Y^i)$  for  $i = 1, 2, \dots, L$ . A MAC of this type is the binary adder channel (BAC) with  $\{0, 1\}$  input alphabets and integer-addition output

$$\underline{Y} = \mathbf{G}_1 \underline{X}_1 + \mathbf{G}_2 \underline{X}_2 \quad (53)$$

where  $\mathbf{G}_1$  and  $\mathbf{G}_2$  are lower-triangular matrices with  $\{0, 1\}$  entries, and where  $\mathbf{G}_1$  has ones on the diagonal.

*Theorem 4:* The capacity region of a MAC with iBM and full feedback and where  $X_{1,i} = f_i(X_2^i, Y^i)$  for all  $i$  is

$$\bigcup \left\{ (R_1, R_2) : \begin{array}{l} 0 \leq R_1, 0 \leq R_2 \\ R_1 \leq I(\mathbf{A}_1^L; Y^L | \mathbf{A}_2^L V) / L \\ R_2 \leq I(\mathbf{A}_2^L; Y^L | \mathbf{A}_1^L V) / L \\ R_1 + R_2 \leq I(\mathbf{A}_1^L \mathbf{A}_2^L; Y^L) / L \end{array} \right\} \quad (54)$$

where the union is over distributions that factor as

$$P(v) \left[ \prod_{k=1}^2 P(\mathbf{a}_k^L | v) 1(x_k^L | \mathbf{a}_k^L, 0y^{L-1}) \right] P(y^L | x_1^L, x_2^L). \quad (55)$$

A cardinality bound on  $V$  is  $|\mathcal{V}| \leq |\mathcal{Y}^L| + 2$ .

*Proof:* The proof mimics that in [20] and is given in Appendix D. ■

*Proposition 1:* An alternative way of writing (56) is

$$\bigcup \left\{ (R_1, R_2) : \begin{array}{l} 0 \leq R_1, 0 \leq R_2 \\ R_1 \leq I(X_1^L \rightarrow Y^L | X_2^L | V) / L \\ R_2 \leq I(X_2^L \rightarrow Y^L | X_1^L | V) / L \\ R_1 + R_2 \leq I(X_1^L X_2^L \rightarrow Y^L) / L \end{array} \right\} \quad (56)$$

where the union is over distributions that factor as

$$P(v) \left[ \prod_{k=1}^2 P(x_k^L | 0y^{L-1} | v) \right] P(y^L | x_1^L, x_2^L). \quad (57)$$

*Proof:* Consider the distribution (55). The chains

$$\mathbf{A}_2^L - V X_2^i Y^{i-1} - Y_i \quad (58)$$

$$\mathbf{A}_1^L \mathbf{A}_2^L - V X_1^i X_2^i Y^{i-1} - Y_i \quad (59)$$

are Markov so that

$$\begin{aligned} & I(\mathbf{A}_1^L; Y^L | \mathbf{A}_2^L V) \\ &= \sum_{i=1}^L H(Y_i | \mathbf{A}_2^L Y^{i-1} X_2^i V) - H(Y_i | \mathbf{A}_1^L \mathbf{A}_2^L Y^{i-1} X_1^i X_2^i V) \\ &= I(X_1^L \rightarrow Y^L | X_2^L | V). \end{aligned} \quad (60)$$

Similarly, the chain  $\mathbf{A}_1^L - V X_1^i Y^{i-1} - Y_i$  is Markov which combined with the Markov chain (59) gives

$$I(\mathbf{A}_2^L; Y^L | \mathbf{A}_1^L V) = I(X_2^L \rightarrow Y^L | X_1^L | V) \quad (61)$$

$$I(\mathbf{A}_1^L \mathbf{A}_2^L; Y^L) = I(X_1^L X_2^L \rightarrow Y^L). \quad (62)$$

The distribution (57) follows from (55). ■

### C. Broadcast Channels

Consider a two-user (three terminal) BC with iBM. We label the transmitter inputs and outputs as  $X^L$  and  $Y^L$ , respectively, and the receiver outputs as  $Y_1^L$  and  $Y_2^L$ . Suppose there are only dedicated messages and no common message. The cut-set bound of Theorem 1 is

$$\bigcup_{P_{\mathbf{A}_L}} \left\{ (R_1, R_2) : \begin{array}{l} 0 \leq R_1 \leq I(\mathbf{A}^L; Y_1^L) / L \\ 0 \leq R_2 \leq I(\mathbf{A}^L; Y_2^L) / L \\ R_1 + R_2 \leq I(\mathbf{A}^L; Y_1^L Y_2^L) / L \end{array} \right\}. \quad (63)$$

An achievable region follows by extending Marton's region as in [1, Lemma 2]: the non-negative rate pair  $(R_1, R_2)$  is achievable if it satisfies

$$\begin{aligned} & LR_1 \leq I(TU_1; Y_1^L) \\ & LR_2 \leq I(TU_2; Y_2^L) \\ & L(R_1 + R_2) \leq \min(I(T; Y_1^L), I(T; Y_2^L)) \\ & \quad + I(U_1; Y_1^L | T) + I(U_2; Y_2^L | T) - I(U_1; U_2 | T) \end{aligned} \quad (64)$$

for some auxiliary random variables  $TU_1U_2$  for which the joint distribution of the random variables factors as

$$P(t, u_1, u_2) P(x^L | 0y^{L-1} | t, u_1, u_2) P(y_1^L, y_2^L | x^L). \quad (65)$$

Marton's region is known to be the same as (63) for  $L = 1$  and *deterministic* broadcast channels. For  $L > 1$ , suppose that  $Y_{1,i}$  and  $Y_{2,i}$  are functions of  $X^i$  for all  $i$ . We may choose

$T = 0$ ,  $U_1 = Y_1^L$ , and  $U_2 = Y_2^L$  without violating the Markov condition (65) and achieve

$$\bigcup_{P_{X^L}} \left\{ (R_1, R_2) : \begin{array}{l} 0 \leq R_1 \leq H(Y_1^L)/L \\ 0 \leq R_2 \leq H(Y_2^L)/L \\ R_1 + R_2 \leq H(Y_1^L Y_2^L)/L \end{array} \right\}. \quad (66)$$

The cut-set region (63) is the same as (66), and therefore (66) is  $\mathcal{C}$ . In fact, feedback does not increase capacity because the transmitter knows, and controls, the channel outputs.

*Remark 24:* The capacity region of a physically degraded BC with two receivers and state known causally at the encoder was derived in [22, Sec. II]. Such channels are NiBMs of length  $L = 2$ , see Remark 16. The cut-set bound of Theorem 1 is loose but the capacity region is achieved by using the coding method described above. In particular, we choose  $U_2$  in (64)-(65) to be a constant and recover the achievability part of Theorem 1 of [22, Sec. II].

#### D. Interference Channels

The cut-set bound is often not so interesting for BCs or interference channels (ICs) with  $L = 1$  because better capacity bounds exist. The same will be true for  $L > 1$ . On the other hand, studying extensions of existing bounds and achievable regions is interesting, e.g., extensions of the Han-Kobayashi region [23] to  $L > 1$ . It may also be interesting to study interference alignment [24], [25] for NiBMs.

### VI. RELAY NETWORKS

Our study of NiBMs was motivated by results on causal relay networks [8] and generalized networks [9]. These networks effectively extend relay networks with delays [7] in the sense that for every relay network with delays there is a causal relay network having the same capacity region. Furthermore, causal relay networks and generalized networks are special NiBMs. This section focuses on relay networks with iBM and applies Theorem 1 to this class of problems.

#### A. Relay Channels

Consider a three-node relay channel (RC) with iBM and source inputs  $X^L$ , relay inputs  $X_1^L$  and outputs  $Y_1^L$ , and destination outputs  $Y^L$ . The RC is a special case of the MAC in Sec. V-A where node 2 (the relay) has no message and the node 1 (the source) has no feedback. A FDG for  $L = 2$  and  $n = 4$  is shown in Fig. 5. The cut-set bound of Theorem 1 is

$$LC \leq \max \min (I(X^L; Y_1^L Y^L | \mathbf{A}_1^L), I(X^L \mathbf{A}_1^L; Y^L)) \quad (67)$$

where the maximization is over  $P_{X^L \mathbf{A}_1^L}$ .

We next list several classic coding strategies [26], [27]. The achievable rates follow by Remark 3 and standard random coding arguments (see [1, Sec. VI]).

- **Decode-forward (DF)** achieves rates  $R$  satisfying

$$LR = \max \min (I(X^L; Y_1^L | \mathbf{A}_1^L), I(X^L \mathbf{A}_1^L; Y^L)) \quad (68)$$

where the maximization is over  $P_{X^L \mathbf{A}_1^L}$  and where the joint distribution factors as

$$P(x^L, \mathbf{a}_1^L) 1(x_1^L \| \mathbf{a}_1^L, 0y_1^{L-1}) P(y_1^L, y^L \| x^L, x_1^L). \quad (69)$$

- **Partial decode-forward (PDF)** achieves  $R$  satisfying

$$LR = \max \min (I(U; Y_1^L | \mathbf{A}_1^L) + I(X^L; Y^L | \mathbf{A}_1^L U), I(X^L \mathbf{A}_1^L; Y^L)) \quad (70)$$

where the maximization is over  $P_{U X^L \mathbf{A}_1^L}$  and where the joint distribution factors as

$$P(u, x^L, \mathbf{a}_1^L) 1(x_1^L \| \mathbf{a}_1^L, 0y_1^{L-1}) P(y_1^L, y^L \| x^L, x_1^L). \quad (71)$$

The rate (70) generalizes [7, Prop. 5].

- **Compress-forward (CF)** achieves  $R$  satisfying

$$LR = \max \min \left( I(X^L; \hat{Y}_1^L Y^L | \mathbf{A}_1^L T), I(X^L \mathbf{A}_1^L; Y^L | T) - I(Y_1^L; \hat{Y}_1^L | X^L \mathbf{A}_1^L Y^L T) \right) \quad (72)$$

where the maximization is over joint distributions that factor as

$$P(t) P(x^L | t) P(\mathbf{a}_1^L | t) 1(x_1^L \| \mathbf{a}_1^L, 0y_1^{L-1}) \cdot P(\hat{y}_1^L | \mathbf{a}_1^L, y_1^L, t) P(y_1^L, y^L \| x^L, x_1^L). \quad (73)$$

*Example 9:* Remark 3 states that we can view the channel as being  $P(y_1^L, y^L | x^L, \mathbf{a}_1^L)$ . The RC is *physically degraded* if the chain

$$X^L - \mathbf{A}_1^L Y_1^L - Y^L$$

is Markov so that  $I(X^L; Y^L | \mathbf{A}_1^L Y_1^L) = 0$ . The DF rate (68) thus matches (67). This capacity result generalizes [7, Prop. 6].

*Remark 25:* Physically degraded RCs with state known causally at the encoder are treated in [22, Sec. III]. Such channels are NiBMs of length  $L = 2$  (see Remark 16) and Theorem 1 gives the converse for Theorem 2 in [22, Sec. IV]. However, these channels fall outside the class of RCs treated in this Section because the source node receives the channel state as “feedback”.

*Example 10:* Suppose the RC is semi-deterministic in the sense that  $Y_{1,i} = f_i(X^i, X_1^i)$  for  $i = 1, 2, \dots, L$ . We may choose  $U = Y_1^L$  and (70) becomes the cut-set bound (67). This capacity result generalizes [7, Prop. 7].

*Example 11:* Suppose the RC is semi-deterministic in the (more general) sense that  $Y_{1,i} = f_i(X^i, X_1^i, Y^i)$  for  $i = 1, 2, \dots, L$ . Consider (72) for which we have

$$I(Y_1^L; \hat{Y}_1^L | X^L \mathbf{A}_1^L Y^L T) = 0. \quad (74)$$

We choose  $T$  as a constant and  $\hat{Y}_1^L = Y_1^L$  so that (72) is the right-hand side of (67) but with independent  $X^L$  and  $\mathbf{A}_1^L$ .

*Example 12:* A special case of Example 11 is where  $Y_{1,i} = f_i(X^i, Y^i)$  and there is a separate channel with iBM and capacity  $R_0$  from the relay to the destination (see [28]). The best  $X^L$  and  $\mathbf{A}_1^L$  are independent so the choice  $\hat{Y}_1^L = Y_1^L$  lets CF achieve the cut-set bound (67).

#### B. Relays without Delay

A relay network with delays is an NiBM in the sense that relabeling time-indexes lets one convert the relay network into an NiBM. We are particularly interested in a relay *without* delay. The corresponding NiBM is a RC with iBM of length  $L = 2$  and the random variables  $X_1, Y_{1,1}, X_{1,2}, Y_2$ , i.e., we have  $\mathcal{X}_{1,1} = \mathcal{Y}_1 = \mathcal{Y}_{1,2} = \mathcal{X}_2 = \{0\}$ . For simplicity, we



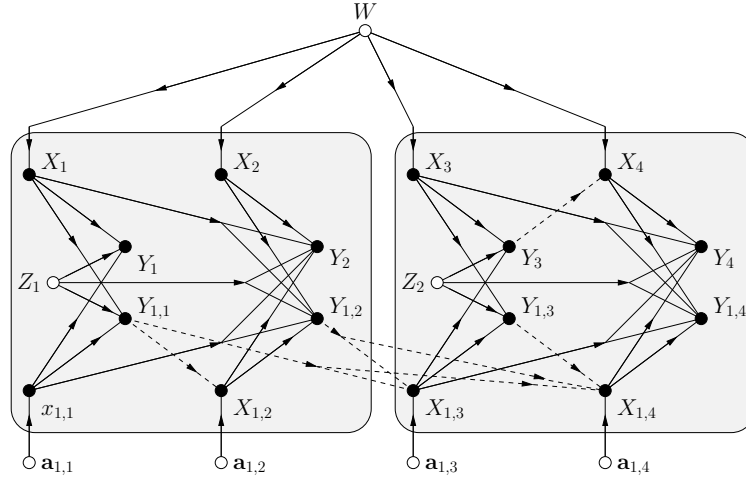


Fig. 5. FDG for a RC with iBM of length  $L = 2$ .

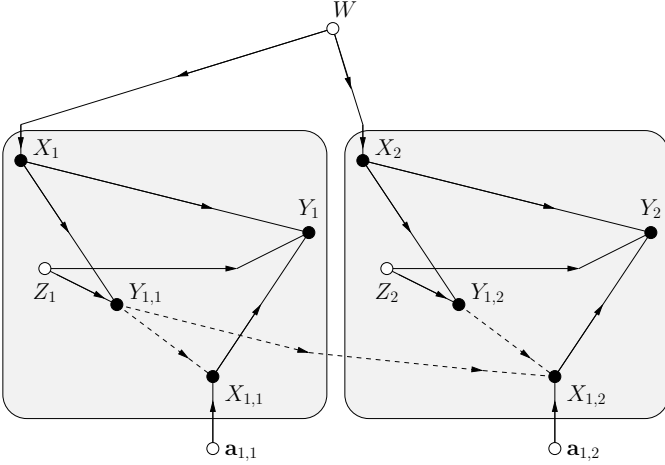


Fig. 6. FDG for a RC when the relay has no delay. The channel is a NiBM of length  $L = 2$ .

rename the variables of interest as  $X, Y_1, X_1, Y$ , respectively, so that the channel is

$$P(y_1^2, y_2^2 \| x^2, x_1^2) = P(y_1 | x) \cdot P(y_2 | x, x_1, y_1). \quad (75)$$

A FDG for 2 channel uses per node is shown in Fig. 6. Observe that this is a subgraph of Fig. 5 where  $L = 2$ ,  $n = 4$ , and where nodes have been relabeled.

The cut-set bound (67) is

$$2C \leq \max \min (I(X; Y_1 Y | \mathbf{A}_1), I(X \mathbf{A}_1; Y)) \quad (76)$$

where the maximization in (76) is over  $P_{X \mathbf{A}_1}$  and  $|\mathbf{A}_1| = |\mathcal{X}_1|^{|\mathcal{Y}_1|}$ . In fact, (76) is simply Willems' bound in [7, p. 3419]. We show in Appendix C that one can choose

$$|\text{supp}(P_{\mathbf{A}_1})| \leq \min(|\mathcal{Y}| + 1, |\mathcal{X}| \cdot |\mathcal{X}_1| + 1). \quad (77)$$

*Remark 26:* The cardinality bound (77) almost agrees with [7, Thm. 2] since  $V$  in [7, Thm. 2] is the code function index. However, the maximization in (76) has a smaller search space in general. To see this, observe that (76) requires optimizing  $P_{X \mathbf{A}_1}$  by considering at most  $N_A = \min(|\mathcal{Y}| + 1, |\mathcal{X}| \cdot |\mathcal{X}_1| + 1)$

out of  $|\mathcal{X}_1|^{|\mathcal{Y}_1|}$  code functions. We must therefore perform at most

$$\binom{|\mathcal{X}_1|^{|\mathcal{Y}_1|}}{N_A}$$

optimizations in  $|X| \cdot N_A - 1$  dimensions. In contrast, [7, Thm. 2] requires optimizing  $P_{X \mathbf{A}_1}$  for  $|\mathcal{X}_1|^{|\mathcal{Y}| \cdot |\mathcal{Y}_1|}$  functions  $f(\cdot) : \mathcal{V} \times \mathcal{Y}_1 \rightarrow \mathcal{X}_1$  where  $|\mathcal{V}|$  is at most  $N_V = |\mathcal{X}| \cdot |\mathcal{X}_1| + 1$ . We thus have at most  $|\mathcal{X}_1|^{N_V \cdot |\mathcal{Y}_1|}$  optimizations in  $|X| \cdot N_V - 1$  dimensions. But we have  $N_A \leq N_V$  and

$$\binom{|\mathcal{X}_1|^{|\mathcal{Y}_1|}}{N_A} \leq |\mathcal{X}_1|^{N_A \cdot |\mathcal{Y}_1|} \leq |\mathcal{X}_1|^{N_V \cdot |\mathcal{Y}_1|} \quad (78)$$

so the optimization in (76) is generally simpler than the optimization in [7, Thm. 2]. This discussion shows that one may as well consider code functions directly rather than introducing auxiliary random variables and auxiliary functions.

*Example 13:* Suppose that  $|\mathcal{X}| = |\mathcal{X}_1| = 2$  and  $|\mathcal{Y}_1| = 4$ . Then (77) states that at most 5 code functions (here code trees) out of 16 need have positive probability. Our search is thus over  $\binom{16}{5} = 4368$  combinations of code trees. In comparison, [7, Thm. 2] requires a search over  $2^{20} \approx 10^6$  mappings  $f(\cdot)$ .

### C. Causal Relay Networks and Generalized Networks

Causal relay networks [8] and generalized networks [9] are NiBMs where every node transmits at most one channel input and receives at most one channel output in each block. For example, the FDG for one block of a causal relay network with  $K = 5$  nodes is shown in Fig. 7. Nodes 1 and 2 are *strictly causal* relays and nodes 3, 4, and 5 are *causal* relays. This relay network is a NiBM of length  $L = 3$ .

In the language of [8], the strictly causal relays are in the set  $\mathcal{N}_1 = \{1, 2\}$  and the causal relays are in  $\mathcal{N}_0 = \{3, 4, 5\}$ . In the language of [9], we have the input and output partitions

$$\begin{aligned} \mathcal{S} &= \{\mathcal{S}_1 = \{1, 2\}, \mathcal{S}_2 = \{3, 4\}, \mathcal{S}_3 = \{5\}\} \\ \mathcal{G} &= \{\mathcal{G}_1 = \{3, 4\}, \mathcal{G}_2 = \{5\}, \mathcal{G}_3 = \{1, 2\}\}. \end{aligned} \quad (79)$$

There are several cut bounds to consider. For example, consider  $\mathcal{S} = \{1, 3\}$  for which [8] uses  $\mathcal{U} = \mathcal{S} \cap \mathcal{N}_0 = \{3\}$ ,

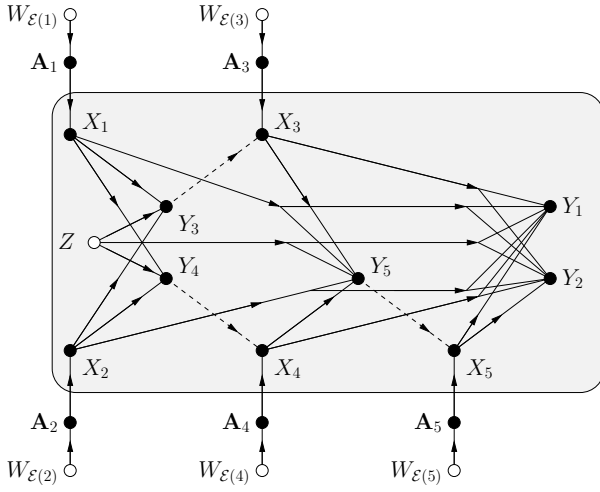


Fig. 7. FDG for a causal relay network with  $K = 5$  nodes and  $n = 3$  channel uses. The network is a NiBM of length  $L = 3$ .

$\mathcal{U}^c = \{4, 5\}$ ,  $\mathcal{V} = \mathcal{S} \cap \mathcal{N}_1 = \{1\}$ ,  $\mathcal{V}^c = \{2\}$ . Using Theorem 1 we have the bounds

$$\begin{aligned}
 3R_{\mathcal{M}(\mathcal{S})} &\stackrel{(a)}{\leq} I(X_1 \mathbf{A}_3; Y_2 Y_4 Y_5 | X_2 \mathbf{A}_4 \mathbf{A}_5) \\
 &\stackrel{(b)}{=} I(X_1; Y_4 | X_2 \mathbf{A}_4 \mathbf{A}_5) \\
 &\quad + I(X_1 \mathbf{A}_3; Y_5 | X_2 X_4 Y_4 \mathbf{A}_4 \mathbf{A}_5) \\
 &\quad + I(X_1 \mathbf{A}_3; Y_2 | X_2 X_4 X_5 Y_4 Y_5 \mathbf{A}_4 \mathbf{A}_5) \\
 &\stackrel{(c)}{\leq} I(X_1; Y_4 | X_2 \mathbf{A}_4 \mathbf{A}_5) \\
 &\quad + I(X_1 X_3 Y_3; Y_5 | X_2 X_4 Y_4 \mathbf{A}_4 \mathbf{A}_5) \\
 &\quad + I(X_1 \mathbf{A}_3; Y_2 | X_2 X_4 X_5 Y_4 Y_5 \mathbf{A}_4 \mathbf{A}_5) \\
 &\stackrel{(d)}{\leq} I(X_1; Y_4 | X_2) \\
 &\quad + I(X_1 X_3 Y_3; Y_5 | X_2 X_4 Y_4) \\
 &\quad + I(X_1 X_3 Y_3; Y_2 | X_2 X_4 X_5 Y_4 Y_5) \quad (80)
 \end{aligned}$$

where (a) is simply (17) and (b) follows by using the chain rule for mutual information and the Markovity in the channel. Step (c) follows because we have added  $Y_3$  to the second mutual information expression and by using the Markovity in the channel. The result is the bound of [8, Thm. 2] when the causal relays do not have their own messages. Step (d) follows similarly and is the bound of [8, Thm. 1] and [9, Thm. 1] when the relays may or may not have their own messages.

Summarizing, we infer that:

- Theorem 1 is at least as good as [8, Thm. 1 and 2] and [9, Thm. 1].
- Example 5 shows that Theorem 1 can be better than [8, Thm. 1] and [9, Thm. 1].
- If the causal relays have no messages then Theorem 1 can be better than [8, Thm. 2] due to inequality (c), see Example 14 below. Furthermore, the auxiliary random variables  $U_k$  in [8, Thm. 2] are not specified to be code functions. The optimization is thus more complex than by using Theorem 1 in general (see Remark 26).

*Example 14:* Consider Fig. 7 with  $\mathcal{X}_k = \mathcal{Y}_k = \{0\}$  for  $k = 2, 4$ , i.e., nodes 2 and 4 are removed from the problem.

Consider  $Y_3 = [X_1, Z]$  where  $\mathcal{X}_1 = \{0, 1\}$  and  $P_Z(0) = P_Z(1) = 1/2$ , and  $Y_5 = Z$ . Suppose there is only one message with rate  $R_{15}$  at node 1 destined for node 5 (so the causal relays at nodes 3 and 5 have no messages). We effectively have a RC with no delay and the capacity is zero because  $X_1 \mathbf{A}_3$  has no influence on  $Y_5$ . For instance, the cut-set bound (17) with  $\mathcal{S} = \{1, 3\}$  gives  $3R_{15} \leq I(X_1 \mathbf{A}_3; Y_5 | \mathbf{A}_5) = 0$ .

Next, consider the cut-set bound of [8, Thm. 2]. There are two cuts to consider without nodes 2 and 4. The cut  $\mathcal{S} = \{1, 3\}$  gives (see (80) after step (c))

$$3R_{15} \leq I(X_1 X_3 Y_3; Y_5 | \mathbf{A}_5) = 1 \quad (81)$$

and the cut  $\mathcal{S} = \{1\}$  gives

$$3R_{15} \leq I(X_1; Y_3 Y_5 | \mathbf{A}_3 \mathbf{A}_5) = H(X_1 | \mathbf{A}_3 \mathbf{A}_5). \quad (82)$$

But we have  $H(X_1 | \mathbf{A}_3 \mathbf{A}_5) = 1$  by choosing  $X_1$  independent of  $\mathbf{A}_3 \mathbf{A}_5$  and  $P_{X_1}(0) = P_{X_1}(1) = 1/2$ . Thus, the cut-set bound of [8, Thm. 2] is loose while Theorem 1 is tight.

*Example 15:* Consider the generalized network called a “BSC with correlated feedback” in [9, Sec. VI]. This network is a two-way channel with iBM of length  $L = 2$  and with binary inputs and outputs

$$\begin{aligned}
 Y_{2,1} &= X_{1,1} \oplus Z \\
 Y_{1,2} &= X_{2,2} \oplus Y_{2,1}
 \end{aligned}$$

where  $P_Z(1) = 1 - P_Z(0) = \epsilon$ . The rate pair  $(R_1, R_2) = (1 - H_2(\epsilon), 1)/2$  is achievable by choosing  $X_{1,1}$  as uniform over  $\{0, 1\}$  and  $X_{2,2} = X'_{2,2} \oplus Y_{2,1}$  where  $X'_{2,2}$  is independent of  $Y_{2,1}$  and uniform over  $\{0, 1\}$ . For the converse, the cut-set bound of Theorem 1 is

$$\bigcup_{P_{X_{1,1} \mathbf{A}_{2,2}}} \left\{ (R_1, R_2) : \begin{aligned} 0 &\leq R_1 \leq I(X_{1,1}; Y_{2,1} | \mathbf{A}_{2,2})/2 \\ 0 &\leq R_2 \leq I(\mathbf{A}_{2,2}; Y_{1,2} | X_{1,1})/2 \end{aligned} \right\} \quad (83)$$

and we have  $I(X_{1,1}; Y_{2,1} | \mathbf{A}_{2,2}) \leq 1 - H_2(\epsilon)$  with equality if  $X_{1,1}$  is uniform and independent of  $\mathbf{A}_{2,2}$ . We further have  $I(\mathbf{A}_{2,2}; Y_{1,2} | X_{1,1}) \leq 1$  since  $Y_{1,2}$  is binary. This shows that Theorem 1 is tight.

Finally, we translate the capacity-achieving strategy into a code tree distribution. We label the branch-pairs of our tree  $\mathbf{A}_{2,2}$  as  $b_0 b_1$  by which we mean that  $X_{2,2} = b_0$  if  $Y_{2,1} = 0$  and  $X_{2,2} = b_1$  if  $Y_{2,1} = 1$ . We choose  $\mathbf{A}_{2,2}$  independent of  $X_{1,1}$  and

$$\begin{aligned}
 P_{\mathbf{A}_{2,2}}(00) &= P_{\mathbf{A}_{2,2}}(11) = 0 \\
 P_{\mathbf{A}_{2,2}}(01) &= P_{\mathbf{A}_{2,2}}(10) = 1/2
 \end{aligned}$$

and compute  $I(\mathbf{A}_{2,2}; Y_{1,2} | X_{1,1}) = 1$ , as desired.

#### D. Digital Network Coding

The final channels we consider are relay networks with iBM. Suppose node 1 multicasts a message of rate  $R$  to sink nodes in the set  $\mathcal{T}$ . The quantize-map-forward (QMF) and noisy network coding (NNC) strategies in [29], [30] generalize to

NiBMs and we call the resulting strategies with code functions *digital network coding* (DNC). DNC achieves  $R$  satisfying

$$LR \leq \min_{k \in \mathcal{S}^c \cap \mathcal{T}} I(\mathbf{A}_{\mathcal{S}}^L; \hat{Y}_{\mathcal{S}^c}^L Y_k | \mathbf{A}_{\mathcal{S}^c}^L T) - I(Y_{\mathcal{S}}^L; \hat{Y}_{\mathcal{S}}^L | \mathbf{A}_{\mathcal{K}}^L \hat{Y}_{\mathcal{S}^c}^L T) \quad (84)$$

for all  $\mathcal{S} \subset \mathcal{K}$  with  $1 \in \mathcal{S}$  and  $\mathcal{S}^c \cap \mathcal{T} \neq \emptyset$ . The  $\mathbf{A}_k^L$ ,  $k = 1, 2, \dots, K$ , are independent and  $\hat{Y}_k^L$  is a noisy function of  $\mathbf{A}_k^L$  and  $Y_k^L$  for all  $k$ .

*Remark 27:* A simple lower bound on the first mutual information expression in (84) is

$$I(\mathbf{A}_{\mathcal{S}}^L; \hat{Y}_{\mathcal{S}^c}^L Y_k | \mathbf{A}_{\mathcal{S}^c}^L T) \geq I(\mathbf{A}_{\mathcal{S}}^L; \hat{Y}_{\mathcal{S}^c}^L | \mathbf{A}_{\mathcal{S}^c}^L T). \quad (85)$$

We use the right-hand side of (85) below because it better matches (17) with  $\hat{Y}_{\mathcal{S}^c}^L$  replacing  $Y_{\mathcal{S}^c}^L$ .

*Example 16:* We extend results of [29], [30]. If the network is deterministic then  $\mathbf{A}_{\mathcal{K}}^L$  determines  $X_{\mathcal{K}}^L Y_{\mathcal{K}}^L$ . We thus have

$$I(Y_{\mathcal{S}}^L; \hat{Y}_{\mathcal{S}}^L | \mathbf{A}_{\mathcal{K}}^L \hat{Y}_{\mathcal{S}^c}^L T) = 0 \quad (86)$$

and can choose  $\hat{Y}_k^L = Y_k^L$  to achieve the cut-set bound but evaluated with independent code functions only. As a result, we obtain the multicast capacity of networks of deterministic point-to-point channels with iBM, for instance. However, DNC will not give the capacity region for all deterministic networks because dependent code functions may increase rates.

### E. DNC for Gaussian Networks

Consider the channel (15) with additive Gaussian noise (AGN), i.e., the  $\underline{Z}_k$  are Gaussian noise vectors and where  $\underline{Z}_{\mathcal{K}}$  has a positive definite covariance matrix. For simplicity, we assume that the  $\underline{Z}_1, \underline{Z}_2, \dots, \underline{Z}_K$  are mutually independent.

Suppose again that node 1 multicasts a message of rate  $R$  to sink nodes in  $\mathcal{T}$ . Let  $\mathcal{S}$  be a cut, i.e.,  $1 \in \mathcal{S}$  and  $\mathcal{S}^c \cap \mathcal{T} \neq \emptyset$ . We use the notation

$$\underline{Y}_{\mathcal{S}^c} = \mathbf{G}_{\mathcal{S}^c \mathcal{S}} \underline{X}_{\mathcal{S}} + \mathbf{G}_{\mathcal{S}^c \mathcal{S}^c} \underline{X}_{\mathcal{S}^c} + \underline{Z}_{\mathcal{S}^c} \quad (87)$$

for the  $|\mathcal{S}^c|$  equations (15) with  $k \in \mathcal{S}^c$ , where  $\mathbf{G}_{\mathcal{U}\mathcal{V}}$  is a  $|\mathcal{U}|L \times |\mathcal{V}|L$  matrix with block-entries  $\mathbf{G}_{kj}$ ,  $k \in \mathcal{U}$ ,  $j \in \mathcal{V}$ .

We begin with the upper bound (25) which we write as

$$\begin{aligned} & h(\mathbf{G}_{\mathcal{S}^c \mathcal{S}} \underline{X}_{\mathcal{S}} + \underline{Z}_{\mathcal{S}^c} | \underline{X}_{\mathcal{S}^c}) - h(\underline{Z}_{\mathcal{S}^c}) \\ & \leq h(\mathbf{G}_{\mathcal{S}^c \mathcal{S}} \underline{X}_{\mathcal{S}} + \underline{Z}_{\mathcal{S}^c}) - h(\underline{Z}_{\mathcal{S}^c}) \\ & \stackrel{(a)}{\leq} \frac{1}{2} \log \frac{|\mathbf{Q}_{\underline{Z}_{\mathcal{S}^c}} + \mathbf{G}_{\mathcal{S}^c \mathcal{S}} \mathbf{Q}_{\underline{X}_{\mathcal{S}}} \mathbf{G}_{\mathcal{S}^c \mathcal{S}}^T|}{|\mathbf{Q}_{\underline{Z}_{\mathcal{S}^c}}|} \end{aligned} \quad (88)$$

where (a) follows by the maximum entropy theorem. The (positive definite) noise covariance matrix has a Cholesky decomposition  $\mathbf{Q}_{\underline{Z}_{\mathcal{S}^c}} = \mathbf{S}_{\underline{Z}_{\mathcal{S}^c}} \mathbf{S}_{\underline{Z}_{\mathcal{S}^c}}^T$  where  $\mathbf{S}_{\underline{Z}_{\mathcal{S}^c}}$  is lower triangular and invertible. We can thus rewrite (88) as

$$I(X_{\mathcal{S}}^L \rightarrow Y_{\mathcal{S}^c}^L | X_{\mathcal{S}^c}^L) \leq \frac{1}{2} \log \left| \mathbf{I}_{\mathcal{S}^c} + \tilde{\mathbf{G}}_{\mathcal{S}^c \mathcal{S}} \mathbf{Q}_{\underline{X}_{\mathcal{S}}} \tilde{\mathbf{G}}_{\mathcal{S}^c \mathcal{S}}^T \right| \quad (89)$$

where  $\mathbf{I}_{\mathcal{U}}$  is the  $|\mathcal{U}|L \times |\mathcal{U}|L$  identity matrix and  $\tilde{\mathbf{G}}_{\mathcal{S}^c \mathcal{S}} = \mathbf{S}_{\underline{Z}_{\mathcal{S}^c}}^{-1} \mathbf{G}_{\mathcal{S}^c \mathcal{S}}$ .

For achievability, we choose  $T$  to be a constant and the code functions (effectively) as code words

$$\mathbf{A}_k^L(\cdot) = X_k^L, \quad k = 1, 2, \dots, K \quad (90)$$

where  $X_k^L$  is Gaussian. We further choose

$$\hat{Y}_k^L = Y_k^L + \hat{Z}_k^L, \quad k = 1, 2, \dots, K \quad (91)$$

where  $\hat{Z}_k^L$  is independent of  $X_k^L Y_k^L$  and has the same statistics as  $Z_k^L$ . Consider the right-hand side of (85) with code words rather than code functions. We have

$$\begin{aligned} & I(X_{\mathcal{S}}^L; \hat{Y}_{\mathcal{S}^c}^L | X_{\mathcal{S}^c}^L) \\ & \stackrel{(a)}{=} h(\mathbf{G}_{\mathcal{S}^c \mathcal{S}} \underline{X}_{\mathcal{S}} + \underline{Z}_{\mathcal{S}^c} + \hat{\underline{Z}}_{\mathcal{S}^c}) - h(\underline{Z}_{\mathcal{S}^c} + \hat{\underline{Z}}_{\mathcal{S}^c}) \\ & \stackrel{(b)}{=} \frac{1}{2} \log \frac{|2\mathbf{Q}_{\underline{Z}_{\mathcal{S}^c}} + \mathbf{G}_{\mathcal{S}^c \mathcal{S}} \mathbf{Q}_{\underline{X}_{\mathcal{S}}} \mathbf{G}_{\mathcal{S}^c \mathcal{S}}^T|}{|2\mathbf{Q}_{\underline{Z}_{\mathcal{S}^c}}|} \\ & = \frac{1}{2} \log \left| \mathbf{I}_{\mathcal{S}^c} + \frac{1}{2} \tilde{\mathbf{G}}_{\mathcal{S}^c \mathcal{S}} \mathbf{Q}_{\underline{X}_{\mathcal{S}}} \tilde{\mathbf{G}}_{\mathcal{S}^c \mathcal{S}}^T \right| \\ & \stackrel{(c)}{\geq} \frac{1}{2} \log \left| \mathbf{I}_{\mathcal{S}^c} + \tilde{\mathbf{G}}_{\mathcal{S}^c \mathcal{S}} \mathbf{Q}_{\underline{X}_{\mathcal{S}}} \tilde{\mathbf{G}}_{\mathcal{S}^c \mathcal{S}}^T \right| - \frac{|\mathcal{S}^c|L}{2} \end{aligned} \quad (92)$$

where (a) is because the  $X_k^L$  are independent, (b) is because the  $X_k^L$  are Gaussian, and (c) follows by using  $|\mathbf{A} + \mathbf{B}| \geq |(\mathbf{A} + \mathbf{B})/2| = |\mathbf{A} + \mathbf{B}|/2^b$  for  $b \times b$  positive definite matrices  $\mathbf{A}$  and  $\mathbf{B}$ . We also have

$$\begin{aligned} I(Y_{\mathcal{S}}^L; \hat{Y}_{\mathcal{S}}^L | X_{\mathcal{K}}^L \hat{Y}_{\mathcal{S}^c}^L) &= I(Z_{\mathcal{S}}^L; Z_{\mathcal{S}}^L + \hat{Z}_{\mathcal{S}}^L | X_{\mathcal{K}}^L \hat{Z}_{\mathcal{S}^c}^L) \\ &= I(Z_{\mathcal{S}}^L; Z_{\mathcal{S}}^L + \hat{Z}_{\mathcal{S}}^L) \\ &= |\mathcal{S}|L/2 \end{aligned} \quad (93)$$

where the last step is because  $\hat{Z}_{\mathcal{S}}^L$  has the same statistics as  $Z_{\mathcal{S}}^L$ . Combining (92) and (93) we find that  $R$  satisfying

$$LR \leq \frac{1}{2} \log \left| \mathbf{I}_{\mathcal{S}^c} + \tilde{\mathbf{G}}_{\mathcal{S}^c \mathcal{S}} \mathbf{Q}_{\underline{X}_{\mathcal{S}}} \tilde{\mathbf{G}}_{\mathcal{S}^c \mathcal{S}}^T \right| - \frac{|\mathcal{K}|L}{2} \quad (94)$$

for all  $\mathcal{S} \subset \mathcal{K}$  with  $1 \in \mathcal{S}$  and  $\mathcal{S}^c \cap \mathcal{T} \neq \emptyset$  are achievable.

It remains to study the first expression on the right-hand side of (94), both without and with independent  $X_k^L$ . Suppose that  $\tilde{\mathbf{G}}_{\mathcal{S}^c \mathcal{S}}$  has the singular value decomposition  $\mathbf{U}^T \mathbf{\Sigma} \mathbf{V}$  so that this expression is

$$\frac{1}{2} \log \left| \mathbf{I}_{\mathcal{S}^c} + \mathbf{\Sigma} \mathbf{V} \mathbf{Q}_{\underline{X}_{\mathcal{S}}} \mathbf{V}^T \mathbf{\Sigma}^T \right|. \quad (95)$$

Suppose there are  $K$  power constraints  $\sum_{i=1}^n \mathbb{E}[X_{k,i}^2]/n \leq P$ ,  $k = 1, 2, \dots, K$  (see (19)). Optimizing over  $\mathbf{Q}_{\underline{X}_{\mathcal{S}}}$  we obtain  $\min(|\mathcal{S}|, |\mathcal{S}^c|) \cdot L$  parallel channels on which we can put at most power  $|\mathcal{S}|P$ . We thus have the capacity upper bound

$$LR \leq \sum_j \frac{1}{2} \log (1 + s_j^2 |\mathcal{S}|P) \quad (96)$$

where the sum is over the parallel channels and the  $s_j$  are the singular values.

For a lower bound we simplify (90) even further and choose  $\mathbf{Q}_{\underline{X}_{\mathcal{K}}} = (P/L) \cdot \mathbf{I}_{\{k\}}$ . The expression (95) becomes

$$\begin{aligned} & \sum_{s_j} \frac{1}{2} \log (1 + s_j^2 (P/L)) \\ & \geq \left[ \sum_{s_j} \frac{1}{2} \log (1 + s_j^2 |\mathcal{S}|P) \right] - \frac{|\mathcal{S}|L}{2} \log (|\mathcal{S}|L). \end{aligned} \quad (97)$$

We thus have the following extension of results in [29], [30] that is interesting for high signal-to-noise ratios.

*Theorem 5:* DNC for scalar, linear, AGN channels, symmetric power constraints, and a multicast session achieves capacity to within

$$|\mathcal{K}|(1 + \log(|\mathcal{K}|L))/2 \text{ bits.} \quad (98)$$

One may derive better results than (98) by using the approach in [30], for example.

#### APPENDIX A PROOF OF CUT-SET BOUND

The bound follows from classic steps and the factorizations (13) and (14). There is one new subtlety, however, namely how to define the random code functions that appear in (17). Fano's inequality states that for  $P_e \rightarrow 0$  we have

$$\begin{aligned} nR_{\mathcal{M}(\mathcal{S})} &\leq I(W_{\mathcal{M}(\mathcal{S})}; \{\hat{W}_{\mathcal{M}(\mathcal{S})}^{(\ell)} : \ell \in \mathcal{S}^c\}) \\ &\stackrel{(a)}{\leq} I(W_{\mathcal{E}(\mathcal{S})}; Y_{\mathcal{S}^c}^n W_{\mathcal{E}(\mathcal{S}^c)}) \\ &\stackrel{(b)}{=} I(W_{\mathcal{E}(\mathcal{S})} \mathbf{A}_{\mathcal{S}}^n; Y_{\mathcal{S}^c}^n | W_{\mathcal{E}(\mathcal{S}^c)} \mathbf{A}_{\mathcal{S}^c}^n) \\ &\stackrel{(c)}{=} I(\mathbf{A}_{\mathcal{S}}^n; Y_{\mathcal{S}^c}^n | \mathbf{A}_{\mathcal{S}^c}^n) \end{aligned} \quad (99)$$

where (a) follows because  $\hat{W}_{\mathcal{M}(\mathcal{S})}$  is a subset of  $\hat{W}_{\mathcal{E}(\mathcal{S})}$  and because  $\{\hat{W}_{\mathcal{M}(\mathcal{S})}^{(\ell)} : \ell \in \mathcal{S}^c\}$  is a function of  $Y_{\mathcal{S}^c}^n$  and  $W_{\mathcal{E}(\mathcal{S}^c)}$ ; (b) follows because the messages are independent and  $\mathbf{A}_{\mathcal{S}}^n$  is a function of the messages at node  $k$ ; and (c) follows because  $W_{\mathcal{E}(\mathcal{S})} - \mathbf{A}_{\mathcal{S}}^L - Y_{\mathcal{S}'}^L$  forms a Markov chain for any  $\mathcal{S}$  and  $\mathcal{S}'$ . Recall that  $n = mL$  for some integer  $m$ . We may thus write

$$\begin{aligned} I(\mathbf{A}_{\mathcal{S}}^n; Y_{\mathcal{S}^c}^n | \mathbf{A}_{\mathcal{S}^c}^n) &\stackrel{(a)}{=} \sum_{i=1}^m I(\mathbf{A}_{\mathcal{S}}^n; Y_{\mathcal{S}^c, i}^L | \mathbf{A}_{\mathcal{S}^c}^n Y_{\mathcal{S}^c}^{(i-1)L}) \\ &\stackrel{(b)}{=} \sum_{i=1}^m I(\mathbf{A}_{\mathcal{S}}^{iL}; Y_{\mathcal{S}^c, i}^L | \mathbf{A}_{\mathcal{S}^c}^{iL} Y_{\mathcal{S}^c}^{(i-1)L}) \\ &\leq \sum_{i=1}^m I(\mathbf{A}_{\mathcal{S}}^{iL} Y_{\mathcal{S}}^{(i-1)L}; Y_{\mathcal{S}^c, i}^L | \mathbf{A}_{\mathcal{S}^c}^{iL} Y_{\mathcal{S}^c}^{(i-1)L}) \end{aligned} \quad (100)$$

where (a) follows by choosing  $Y_{k, i}^L$  to be the channel output of node  $k$  from time  $(i-1)L + 1$  to time  $iL$ , and where (b) follows by Markovity.

We now define  $\bar{\mathbf{A}}_{k, i}^L$  to be the sub-function of  $\mathbf{A}_k^{iL}$  of depth  $L$  that corresponds to the channel output  $Y_k^{(i-1)L}$ . We then have

$$\begin{aligned} I(\mathbf{A}_{\mathcal{S}}^{iL} Y_{\mathcal{S}}^{(i-1)L}; Y_{\mathcal{S}^c, i}^L | \mathbf{A}_{\mathcal{S}^c}^{iL} Y_{\mathcal{S}^c}^{(i-1)L}) \\ &\stackrel{(a)}{=} H(Y_{\mathcal{S}^c, i}^L | \bar{\mathbf{A}}_{\mathcal{S}^c, i}^L \mathbf{A}_{\mathcal{S}^c}^{iL} Y_{\mathcal{S}^c}^{(i-1)L}) - H(Y_{\mathcal{S}^c, i}^L | \bar{\mathbf{A}}_{\mathcal{K}, i}^L \mathbf{A}_{\mathcal{K}}^{iL} Y_{\mathcal{K}}^{(i-1)L}) \\ &\leq H(Y_{\mathcal{S}^c, i}^L | \bar{\mathbf{A}}_{\mathcal{S}^c, i}^L) - H(Y_{\mathcal{S}^c, i}^L | \bar{\mathbf{A}}_{\mathcal{K}, i}^L \mathbf{A}_{\mathcal{K}}^{iL} Y_{\mathcal{K}}^{(i-1)L}) \\ &\stackrel{(b)}{=} H(Y_{\mathcal{S}^c, i}^L | \bar{\mathbf{A}}_{\mathcal{S}^c, i}^L) - H(Y_{\mathcal{S}^c, i}^L | \bar{\mathbf{A}}_{\mathcal{K}, i}^L) \\ &= I(\bar{\mathbf{A}}_{\mathcal{S}, i}^L; Y_{\mathcal{S}^c, i}^L | \bar{\mathbf{A}}_{\mathcal{S}^c, i}^L) \end{aligned} \quad (101)$$

where (a) follows because  $\bar{\mathbf{A}}_{k, i}^L$  is a function of  $\mathbf{A}_k^{iL} Y_k^{(i-1)L}$  and (b) follows because  $\mathbf{A}_{\mathcal{K}}^{iL} Y_{\mathcal{K}}^{(i-1)L} - \bar{\mathbf{A}}_{\mathcal{K}, i}^L - Y_{\mathcal{S}^c, i}^L$  forms a Markov chain (this step is crucial because it permits  $L$ -letterization).

The remaining steps follow in the usual way because the  $\bar{\mathbf{A}}_{\mathcal{K}}^L$ -to- $Y_{\mathcal{K}}^L$  channel does not depend on the block index  $i$ . More precisely, we have

$$\begin{aligned} P(y_{\mathcal{K}, i}^L | \bar{\mathbf{a}}_{\mathcal{K}, i}^L) &= P_{Y_{\mathcal{K}}^L | \mathbf{A}_{\mathcal{K}}^L}(y_{\mathcal{K}, i}^L | \bar{\mathbf{a}}_{\mathcal{K}, i}^L) \\ &= \left[ \prod_{k=1}^K 1(x_k^L | \bar{\mathbf{a}}_{k, i}^L, 0 y_{k, i}^{L-1}) \right] P_{Y_{\mathcal{K}}^L | X_{\mathcal{K}}^L}(y_{\mathcal{K}, i}^L | x_{\mathcal{K}}^L) \end{aligned} \quad (102)$$

where  $P_{Y_{\mathcal{K}}^L | \mathbf{A}_{\mathcal{K}}^L}$  refers to the first  $L$  channel uses. Inserting (101) into (100), we have

$$\begin{aligned} I(\mathbf{A}_{\mathcal{S}}^n; Y_{\mathcal{S}^c}^n | \mathbf{A}_{\mathcal{S}^c}^n) &\leq \sum_{i=1}^m I(\bar{\mathbf{A}}_{\mathcal{S}, i}^L; Y_{\mathcal{S}^c, i}^L | \bar{\mathbf{A}}_{\mathcal{S}^c, i}^L) \\ &= m I(\bar{\mathbf{A}}_{\mathcal{S}, T}^L; Y_{\mathcal{S}^c, T}^L | \bar{\mathbf{A}}_{\mathcal{S}^c, T}^L) \\ &\stackrel{(a)}{\leq} m I(\bar{\mathbf{A}}_{\mathcal{S}, T}^L; Y_{\mathcal{S}^c, T}^L | \bar{\mathbf{A}}_{\mathcal{S}^c, T}^L) \end{aligned} \quad (103)$$

where  $T$  takes on the value  $i$ ,  $i = 1, 2, \dots, m$ , with probability  $1/m$ , and where (a) follows because  $T - \bar{\mathbf{A}}_{\mathcal{K}, T}^L - Y_{\mathcal{K}, T}^L$  forms a Markov chain. Inserting (103) into (99), we have

$$L \cdot R_{\mathcal{M}(\mathcal{S})} \leq I(\bar{\mathbf{A}}_{\mathcal{S}, T}^L; Y_{\mathcal{S}^c, T}^L | \bar{\mathbf{A}}_{\mathcal{S}^c, T}^L) \quad (104)$$

where the joint distribution of the random variables factors as

$$P(\bar{\mathbf{a}}_{\mathcal{K}, T}^L) P_{Y_{\mathcal{K}}^L | \mathbf{A}_{\mathcal{K}}^L}(y_{\mathcal{K}, T}^L | \bar{\mathbf{a}}_{\mathcal{K}, T}^L) \quad (105)$$

and where the second term in (105) is computed using (102) (this fact is crucial because it permits the factorization (18)).

*Remark 28:* Consider the case  $n \neq mL$  for which we may as well consider  $n = mL + L'$  where  $0 < L' < L$ . The sum in (103) will then be increased by a term of the form

$$I(\bar{\mathbf{A}}_{\mathcal{S}, m+1}^{L'}; Y_{\mathcal{S}^c, m+1}^{L'} | \bar{\mathbf{A}}_{\mathcal{S}^c, m+1}^{L'}) \quad (106)$$

where the code functions have depth  $L'$ . The term (106) could be larger than the right-hand side of (104). However, if (106) is bounded and  $m$  is large then the capacity is effectively limited by (104).

*Remark 29:* Consider the  $j$ th cost constraint in (19). We may rewrite (19) as

$$\begin{aligned} \frac{1}{L} \sum_{\ell=1}^L \frac{1}{m} \sum_{i=1}^m \mathbb{E} [s_j (X_{\mathcal{K}, (m-1)L+\ell}, Y_{\mathcal{K}, (m-1)L+\ell})] \\ = \frac{1}{L} \sum_{\ell=1}^L \mathbb{E} [s_j (X_{\mathcal{K}, (T-1)L+\ell}, Y_{\mathcal{K}, (T-1)L+\ell})] \leq S_j \end{aligned} \quad (107)$$

and the inequality in (107) is the  $j$ th inequality in (20).

#### APPENDIX B CARDINALITY BOUNDS FOR POINT-TO-POINT CHANNELS

Consider a point-to-point channel with NiBM. We write

$$P(y^L) = \sum_{\mathbf{a}^L} P(\mathbf{a}^L) P(y^L | \mathbf{a}^L) \quad (108)$$

$$H(Y^L | \mathbf{A}^L) = \sum_{\mathbf{a}^L} P(\mathbf{a}^L) H(Y^L | \mathbf{A}^L = \mathbf{a}^L) \quad (109)$$

where  $P(y^L | \mathbf{a}^L)$  and  $H(Y^L | \mathbf{A}^L = \mathbf{a}^L)$  are determined by the channel  $P(y^L | x^L)$ . Equations (108) and (109) imply that  $P(y^L)$  and  $H(Y^L | \mathbf{A}^L)$  are convex combinations of  $P(\mathbf{a}^L)$ .

Furthermore, if we fix  $P_{Y^L}(\cdot)$  and  $H(Y^L|\mathbf{A}^L)$  then we have fixed  $I(\mathbf{A}^L; Y^L)$ . We can therefore focus on  $|\mathcal{Y}^L|$  constraints and [14, Lemma 3.4] guarantees that we need only  $|\mathcal{Y}^L|$  non-zero values of  $P(\mathbf{a}^L)$ .

Similarly, observe that

$$P(y^L) = \sum_{x^L, \tilde{y}^L} P(x^L \| 0\tilde{y}^{L-1}) P(\tilde{y}^L, y^L \| x^L) \quad (110)$$

so that if we fix  $P(x^L \| 0\tilde{y}^{L-1})$  then we have fixed  $P(y^L)$ . Our approach will be to replace  $|\mathcal{Y}^L| - 1$  constraints of the form (108) with (hopefully fewer) constraints to fix  $P(x^L \| 0\tilde{y}^{L-1})$ .

We proceed by induction. We may fix  $P(x_1)$  with  $|\mathcal{X}_1| - 1$  constraints of the form

$$P(x_1) = \sum_{\mathbf{a}^L} P(\mathbf{a}^L) P(x_1 | \mathbf{a}^L) \quad (111)$$

since  $P(x_1 | \mathbf{a}^L)$  is a fixed function. This fixes  $P(x_1, \tilde{y}_1)$  because the channel specifies  $P(\tilde{y}_1 | x_1)$ . Now suppose that  $P(x^{i-1}, \tilde{y}^{i-1})$  is fixed and write

$$P(x_i | x^{i-1}, \tilde{y}^{i-1}) = \sum_{\mathbf{a}^L} P(\mathbf{a}^L) \frac{P(x^i, \tilde{y}^{i-1} | \mathbf{a}^L)}{P(x^{i-1}, \tilde{y}^{i-1})} \quad (112)$$

where  $P(x^i, \tilde{y}^{i-1} | \mathbf{a}^L)$  is fixed because  $\mathbf{a}^L$  is in the condition- ing. We must thus define

$$|\mathcal{X}^{i-1}| \cdot |\mathcal{Y}^{i-1}| \cdot (|\mathcal{X}_i| - 1) \quad (113)$$

constraints of the form (112) to fix  $P(x_i | x^{i-1}, \tilde{y}^{i-1})$  for all its arguments. This in turn fixes  $P(x_i, \tilde{y}_i | x^{i-1}, \tilde{y}^{i-1})$  because the channel specifies  $P(\tilde{y}_i | x^i, \tilde{y}^{i-1})$ . We thus find that  $P(x^i, \tilde{y}^i)$  is fixed which completes the induction step. Collecting all the constraints including (109) we have

$$|\mathcal{X}_1| + \sum_{i=2}^L |\mathcal{X}^{i-1}| \cdot |\tilde{\mathcal{Y}}^{i-1}| \cdot (|\mathcal{X}_i| - 1) \quad (114)$$

constraints in total. This number may be less than  $|\mathcal{Y}^L|$ , e.g., if one of the  $L$  channel outputs is continuous.

#### APPENDIX C

##### CARDINALITY BOUNDS FOR RELAYS WITHOUT DELAY

Consider an RC without delay. Suppose  $P(x | \mathbf{a}_1)$  is fixed which fixes  $P(x, x_1, y, y_1 | \mathbf{a}_1)$  because the channel fixes  $P(y_1 | x)$  and  $P(y | x, x_1, y_1)$ , and  $\mathbf{a}_1$  fixes  $P(x_1 | \mathbf{a}_1, y_1)$  due to (1). We have

$$P(y) = \sum_{\mathbf{a}_1} P(\mathbf{a}_1) P(y | \mathbf{a}_1) \quad (115)$$

$$H(Y | X \mathbf{A}_1) = \sum_{\mathbf{a}_1} P(\mathbf{a}_1) H(Y | X, \mathbf{A}_1 = \mathbf{a}_1) \quad (116)$$

$$I(X; Y Y_1 | \mathbf{A}_1) = \sum_{\mathbf{a}_1} P(\mathbf{a}_1) I(X; Y Y_1 | \mathbf{A}_1 = \mathbf{a}_1) \quad (117)$$

where  $P(y | \mathbf{a}_1)$ ,  $H(Y | X, \mathbf{A}_1 = \mathbf{a}_1)$ , and  $I(X; Y Y_1 | \mathbf{A}_1 = \mathbf{a}_1)$  are all fixed quantities. Finally, if we fix  $P_Y(\cdot)$  and  $H(Y | X \mathbf{A}_1)$  then we have fixed  $I(X \mathbf{A}_1; Y)$ . We thus have  $|\mathcal{Y}| + 1$  constraints in total.

Next, we note that

$$P(y) = \sum_{x, x_1, y_1} P(x, x_1) P(y_1 | x) P(y | x, x_1, y_1) \quad (118)$$

so that if we fix  $P(x, x_1)$  then we have fixed  $P(y)$ . We proceed by writing

$$P(x, x_1) = \sum_{\mathbf{a}_1} P(\mathbf{a}_1) P(x, x_1 | \mathbf{a}_1) \quad (119)$$

which gives us  $|\mathcal{X}| \cdot |\mathcal{X}_1| - 1$  constraints instead of the  $|\mathcal{Y}| - 1$  before. Together with (116) and (117) we arrive at  $|\mathcal{X}| \cdot |\mathcal{X}_1| + 1$  constraints in total.

#### APPENDIX D

##### CONVERSE FOR A CLASS OF MACS WITH FEEDBACK

Let  $V_i = X_1^{(i-1)L} Y^{(i-1)L}$  for  $i = 1, 2, \dots, m$ . Fano's inequality and the independence of the messages gives

$$\begin{aligned} nR_1 &\leq I(W_1; Y^n | W_2) \\ &= I(\mathbf{A}_1^n; Y^n | \mathbf{A}_2^n) \\ &= \sum_{i=1}^m H(Y_i^L | \mathbf{A}_2^{iL} Y^{(i-1)L}) - H(Y_i^L | \mathbf{A}_1^{iL} \mathbf{A}_2^{iL} Y^{(i-1)L}) \\ &\stackrel{(a)}{=} \sum_{i=1}^m H(Y_i^L | \mathbf{A}_2^{iL} V_i) - H(Y_i^L | \mathbf{A}_1^{iL} \mathbf{A}_2^{iL} V_i) \\ &\stackrel{(b)}{\leq} mI(\bar{\mathbf{A}}_{1,T}^L; Y_T^L | \bar{\mathbf{A}}_{2,T}^L V_T) \\ &\stackrel{(c)}{\leq} mI(\bar{\mathbf{A}}_{1,T}^L; Y_T^L | \bar{\mathbf{A}}_{2,T}^L V_T) \end{aligned} \quad (120)$$

where (a) follows because  $\mathbf{A}_2^i Y^{i-1}$  defines  $X_2^i$  and therefore also  $X_1^{i-1}$ . Step (b) follows by using  $T$  as our time-sharing random variable,  $\bar{\mathbf{A}}_{k,i}^L$  as in Appendix A, and similar steps as in (101); step (c) follows because

$$T - V_T \bar{\mathbf{A}}_{1,T}^L \bar{\mathbf{A}}_{2,T}^L - Y_T^L$$

forms a Markov chain. The chains

$$\begin{aligned} T - \bar{\mathbf{A}}_{1,T}^L \bar{\mathbf{A}}_{2,T}^L - Y_T^L \\ \bar{\mathbf{A}}_{1,T}^L - V_T - \bar{\mathbf{A}}_{2,T}^L \end{aligned}$$

are also Markov.

By symmetry, we have a similar bound for  $nR_2$ . The corresponding sum-rate bound is

$$\begin{aligned} n(R_1 + R_2) &\leq I(W_1 W_2; Y^n) \\ &= I(\mathbf{A}_1^n \mathbf{A}_2^n; Y^n) \\ &\leq \sum_{i=1}^m H(Y_i^L) - H(Y_i^L | \mathbf{A}_1^{iL} \mathbf{A}_2^{iL} V_i) \\ &= mI(\bar{\mathbf{A}}_{1,T}^L \bar{\mathbf{A}}_{2,T}^L; Y_T^L | T) \\ &\leq mI(\bar{\mathbf{A}}_{1,T}^L \bar{\mathbf{A}}_{2,T}^L; Y_T^L). \end{aligned} \quad (121)$$

Collecting the bounds, we arrive at the region of Theorem 4. The cardinality bound follows by using similar steps as in Appendices B and C, see also [31, App. B].

## ACKNOWLEDGMENTS

The results reported here were motivated by the paper [9] by S. Fong and R. Yeung. I am grateful to these authors for sending me of an early version of their work. I am also grateful to S.-Y. Chung and H. Permuter for their detailed and constructive criticisms of the document.

## REFERENCES

- [1] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.
- [2] C. E. Shannon, "Two-way communication channels," in *Proc. 4th Berkeley Symp. on Mathematical Statistics and Probability*, J. Neyman, Ed., Berkeley, CA, 1961, vol. 1, pp. 611–644, Univ. Calif. Press, Reprinted in *Claude Elwood Shannon: Collected Papers*, pp. 351–384, (N.J.A. Sloane and A.D. Wyner, eds.) Piscataway: IEEE Press, 1993.
- [3] R.E. Blahut, *Principles and Practice of Information Theory*, Addison-Wesley, Reading, Massachusetts, 1987.
- [4] G. Kramer, "Topics in multi-user information theory," *Foundations and Trends in Networking*, vol. 4, no. 4-5, pp. 265–444, 2007.
- [5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, 2nd edition, 2006.
- [6] A. El Gamal and Y.-H. Kim, *Network Information Theory*, Cambridge Univ. Press, 2011.
- [7] A. El Gamal, N. Hassanpour, and J. Mammen, "Relay networks with delays," *IEEE Trans. Inf. Theory*, vol. 53, pp. 3413–3431, Oct. 2007.
- [8] I.-J. Baik and S.-Y. Chung, "Causal relay networks and new cut-set bounds," in *Proc. 49th Annual Allerton Conf. on Commun., Control, and Computing*, Monticello, IL, Sept. 2011, see also <http://arxiv.org/abs/1207.4587>.
- [9] S. L. Fong and R. W. Yeung, "Cut-set bounds for generalized networks," *IEEE Trans. Inf. Theory*, Aug. 2012, submitted.
- [10] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, pp. 221–238, Sept. 1956, Reprinted in *Claude Elwood Shannon: Collected Papers*, pp. 221–238, (N.J.A. Sloane and A.D. Wyner, eds.) Piscataway: IEEE Press, 1993.
- [11] R. E. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, pp. 14–20, Jan. 1972.
- [12] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, pp. 460–473, July 1972.
- [13] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, New York, 1968.
- [14] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Channels*, Akadémiai Kiadó, Budapest, 1981.
- [15] R. J. McEliece and W. E. Stark, "Channels with block interference," *IEEE Trans. Inf. Theory*, vol. 30, no. 1, pp. 44–53, Jan. 1984.
- [16] L. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic consideration for cellular mobile radio," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 359–378, May 1994.
- [17] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, pp. 289–293, Oct. 1958, Reprinted in *Claude Elwood Shannon: Collected Papers*, pp. 273–278, (N.J.A. Sloane and A.D. Wyner, eds.) Piscataway: IEEE Press, 1993.
- [18] H. Farmanbar and A. K. Khandani, "Precoding for the AWGN channel with discrete interference," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4019–4032, Sept. 2009.
- [19] T. Weissman, "Capacity of channels with action-dependent states," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5396–5411, Nov. 2010.
- [20] F. M. J. Willems, "The feedback capacity region of a class of discrete memoryless multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 28, no. 1, pp. 93–95, Jan. 1982.
- [21] T. M. Cover and C. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 27, no. 3, pp. 292–298, May 1981.
- [22] S. Sigurjonsson and Y. H. Kim, "On multiple user channels with state information at the transmitters," in *Proc. 2005 IEEE Int. Symp. Inform. Theory*, Adelaide, Australia, Sept. 2005, pp. 72–76.
- [23] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 49–60, Jan. 1981.
- [24] V. R. Cadambe and S.-A. Jafar, "Interference alignment and degrees of freedom of the  $k$ -user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.
- [25] M. A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over MIMO X channels: interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.
- [26] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.
- [27] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sept. 2005.
- [28] Y.-H. Kim, "Capacity of a class of deterministic relay channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1328–1329, Mar. 2008.
- [29] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: a deterministic approach," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1872–1905, Apr. 2011.
- [30] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.
- [31] F. M. J. Willems and E. C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 313–327, May 1985.